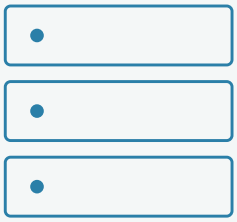
LESSON BOOK



Data Center Systems

Power, heat, cooling, measurement, control and efficiency

Twenty-four lessons · 575 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Data Center Systems

Power, heat, cooling, measurement and control as one physical system

A lesson sequence, written up as a book

Ricardo Marquez

October 5, 2026

How to read this book

This book treats a data center as a physical system rather than as a room of computers. Each chapter expands one lesson of a sequence that follows energy through the facility,

$$\text{electricity} \rightarrow \text{IT equipment} \rightarrow \text{heat} \rightarrow \text{air and water} \rightarrow \text{environment},$$

and then asks what the facility measures of that chain, how it acts on it, and how its efficiency is reported. Chapters 12 to 20 turn to operation under uncertainty: reliability, diagnosis, state estimation, identifiability, capacity as a risk, what-if analysis, the choice of the next measurement, and optimization. Chapters 21 to 24 close the book: the facility as a flexible load on the grid, the choice of model fidelity, the data that are publicly available, and a synthesis. The emphasis throughout is on physical modeling, operation, uncertainty, diagnosis and control, not on a catalogue of equipment.

Three questions recur in every chapter and organize the book: what is *measured*, what is *calculated* from measurements, and what must be *inferred*. A rack inlet temperature is measured; the heat a rack generates is calculated from its electrical power; a recirculation path or a degraded cooling coil is inferred. The numbers used in the examples are illustrative, chosen to make each relation concrete; they are not data from a particular facility.

Where the field uses a shorthand term, a *Jargon* box expands it once and the text then uses the expanded meaning. Every symbol is defined in the notation chapter that follows and is used with the same meaning in every chapter. Each chapter closes with a statement of what to retain, exercises, and a short section connecting the chapter to the TERRA framework, a graph-based way of modeling physical systems as conservation laws on nodes with flows on edges, with probabilistic inference of the states and parameters that the measurements do not give directly.

Contents

How to read this book **ii**

Notation and vocabulary **ix**

1 System architecture **1**
1.1 The facility as an energy-conversion system 1
1.2 The electrical and IT subsystems 4
1.3 The path of heat 5
1.4 Three coupled networks and the control loop 10
1.5 Redundancy and measurement 13
1.6 Diagnosis and the system picture 16
1.7 What to retain 20
1.8 Exercises 20
1.9 Connection to the TERRA framework 21

2 IT loads and computing demand **23**
2.1 Server power 23
2.2 Rack power and rack density 27
2.3 From computing demand to load 29
2.4 Flexibility of the load 31
2.5 Time variation and facility load 34
2.6 Telemetry, diagnosis and state space 37
2.7 What to retain 40
2.8 Exercises 40
2.9 Connection to the TERRA framework 43

3 Electrical power distribution **44**
3.1 Continuity of power 44
3.2 From the utility to the UPS 46
3.3 From the UPS to the server 50
3.4 Redundancy and contingency capacity 53
3.5 Losses and the facility power balance 56
3.6 Topology, state and reliability 59
3.7 What to retain 63
3.8 Exercises 63
3.9 Connection to the TERRA framework 65

4 UPS, batteries and backup power **67**
4.1 Two roles in one transition 67
4.2 Sizing the battery 70
4.3 The generator and the transfer 73

4.4 Load priority and the thermal buffer . . . 76
4.5 Battery health and runtime . . . 78
4.6 Dependence, diagnosis and dual use . . . 81
4.7 What to retain . . . 87
4.8 Exercises . . . 87
4.9 Connection to the TERRA framework . . . 89

5 Heat generation and thermal balance 90
5.1 Conservation of energy . . . 90
5.2 The rack and the air stream . . . 93
5.3 The room and the water loop . . . 95
5.4 Resistance, capacitance and time . . . 98
5.5 Spatial structure and mixing . . . 101
5.6 Closing the balance and inference . . . 105
5.7 What to retain . . . 108
5.8 Exercises . . . 110
5.9 Connection to the TERRA framework . . . 110

6 Airflow and rack cooling 112
6.1 Airflow required by a rack . . . 112
6.2 Separation and mixing . . . 114
6.3 Pressure, resistance and fans . . . 118
6.4 Distribution, containment and hot spots . . . 122
6.5 Measuring and inferring airflow . . . 125
6.6 Network models, CFD and control . . . 127
6.7 What to retain . . . 132
6.8 Exercises . . . 133
6.9 Connection to the TERRA framework . . . 134

7 Cooling equipment 135
7.1 The heat-rejection chain . . . 135
7.2 The coil and the chilled-water loop . . . 137
7.3 Pumps and the chiller . . . 140
7.4 Heat rejection to the atmosphere . . . 144
7.5 The plant as a network with capacities . . . 147
7.6 Measurement, diagnosis and control . . . 151
7.7 What to retain . . . 155
7.8 Exercises . . . 156
7.9 Connection to the TERRA framework . . . 157

8 Thermal dynamics 159
8.1 The dynamic heat balance . . . 159
8.2 The first-order thermal model . . . 162
8.3 Memory, multiple masses and storage . . . 164
8.4 Propagation, delays and diagnosis . . . 168
8.5 Control interaction and dynamic capacity . . . 171
8.6 State space, estimation and identification . . . 174
8.7 What to retain . . . 178
8.8 Exercises . . . 178
8.9 Connection to the TERRA framework . . . 180

9 Sensors and instrumentation **182**
9.1 The measurement equation . . . 182
9.2 What is measured . . . 184
9.3 Time and error . . . 190
9.4 Checks, residuals and observability . . . 194
9.5 Design, placement and estimated quantities . . . 196
9.6 From raw sensors to risk . . . 199
9.7 What to retain . . . 203
9.8 Exercises . . . 203
9.9 Connection to the TERRA framework . . . 205

10 Controls **207**
10.1 The control problem . . . 207
10.2 Feedback controllers . . . 209
10.3 The loops of a cooling plant . . . 215
10.4 Predictive and coordinated control . . . 218
10.5 Control under uncertainty . . . 222
10.6 Diagnosis-aware and supervisory control . . . 225
10.7 What to retain . . . 229
10.8 Exercises . . . 229
10.9 Connection to the TERRA framework . . . 230

11 PUE and energy efficiency **232**
11.1 Power usage effectiveness . . . 232
11.2 What the ratio does not measure . . . 234
11.3 Why PUE varies . . . 237
11.4 Decomposition and local problems . . . 240
11.5 PUE in time . . . 244
11.6 Capacity, evidence and decisions . . . 247
11.7 What to retain . . . 251
11.8 Exercises . . . 252
11.9 Connection to the TERRA framework . . . 253

12 Redundancy and reliability **255**
12.1 Redundancy levels and layers . . . 255
12.2 Dependence and single points of failure . . . 258
12.3 Availability, maintenance and capacity states . . . 262
12.4 Contingency analysis in time . . . 265
12.5 Event trees and fault trees . . . 268
12.6 Operational risk . . . 271
12.7 What to retain . . . 275
12.8 Exercises . . . 275
12.9 Connection to the TERRA framework . . . 276

13 Failure modes and diagnostics **278**
13.1 Symptoms, causes and hypotheses . . . 278
13.2 Air-side failure modes . . . 280
13.3 Water-side and plant failure modes . . . 283
13.4 Sensor, topology and electrical faults . . . 286
13.5 Residual-based diagnosis and ranking . . . 290
13.6 From diagnosis to action . . . 296
13.7 What to retain . . . 299

13.8 Exercises . . . 300
13.9 Connection to the TERRA framework . . . 300

14 State estimation **302**
14.1 Measurements and the physical state . . . 302
14.2 Dynamic state estimation . . . 307
14.3 Static estimation and residuals . . . 311
14.4 Estimated quantities in a data center . . . 314
14.5 Observability and posterior uncertainty . . . 317
14.6 Use and limits of state estimation . . . 319
14.7 What to retain . . . 324
14.8 Exercises . . . 324
14.9 Connection to the TERRA framework . . . 325

15 Uncertainty and identifiability **327**
15.1 Uncertainty and identifiability . . . 327
15.2 Parameter combinations and posterior geometry . . . 331
15.3 Ambiguity in diagnosis . . . 336
15.4 Priors, adequacy and model structure . . . 338
15.5 Capability and the limits of inference . . . 341
15.6 Identifiability relative to the decision . . . 343
15.7 What to retain . . . 347
15.8 Exercises . . . 348
15.9 Connection to the TERRA framework . . . 348

16 Probabilistic diagnosis **350**
16.1 Diagnosis as inference . . . 350
16.2 Bayes' rule on a worked case . . . 351
16.3 Likelihoods from physical models . . . 354
16.4 Identifiability, confidence and scale . . . 358
16.5 Time, normal operation and model error . . . 361
16.6 From posterior to action . . . 364
16.7 What to retain . . . 369
16.8 Exercises . . . 369
16.9 Connection to the TERRA framework . . . 370

17 Capacity and risk **372**
17.1 Capacity as a system property . . . 372
17.2 State-dependent and secure capacity . . . 374
17.3 Local and constrained capacity . . . 376
17.4 Probabilistic capacity . . . 379
17.5 Capacity margin in time . . . 383
17.6 Bottlenecks and capacity planning . . . 387
17.7 What to retain . . . 391
17.8 Exercises . . . 391
17.9 Connection to the TERRA framework . . . 393

18 What-if analysis **395**
18.1 The what-if question . . . 395
18.2 Load scenarios . . . 398
18.3 Failure and weather scenarios . . . 401
18.4 Uncertainty in scenarios . . . 405

18.5 Actions and sensitivity . . . 409
18.6 Model credibility . . . 414
18.7 What to retain . . . 416
18.8 Exercises . . . 417
18.9 Connection to the TERRA framework . . . 418

19 Next-best measurement **420**
19.1 An ambiguous diagnosis . . . 420
19.2 Information gain . . . 423
19.3 Decision value . . . 427
19.4 Active tests and experiment design . . . 429
19.5 Sequential and question-specific measurement . . . 433
19.6 Limits and reporting . . . 436
19.7 What to retain . . . 438
19.8 Exercises . . . 440
19.9 Connection to the TERRA framework . . . 440

20 Data center optimization **442**
20.1 The optimization problem . . . 442
20.2 Constraints and objectives . . . 446
20.3 Workload, sites and storage . . . 448
20.4 Optimization under uncertainty . . . 452
20.5 Dynamics, discrete decisions, contingencies . . . 456
20.6 Optimization in operation . . . 460
20.7 What to retain . . . 462
20.8 Exercises . . . 463
20.9 Connection to the TERRA framework . . . 464

21 Grid interaction and flexible load **466**
21.1 Flexibility of a facility . . . 466
21.2 Electrical and IT resources . . . 470
21.3 Thermal, generation and portfolio . . . 473
21.4 Baselines and forecasts of flexibility . . . 477
21.5 Committing flexibility . . . 479
21.6 Reserve, signals and reporting . . . 483
21.7 What to retain . . . 487
21.8 Exercises . . . 487
21.9 Connection to the TERRA framework . . . 489

22 Digital twins, CFD and reduced models **490**
22.1 Three kinds of model . . . 490
22.2 Computational fluid dynamics . . . 492
22.3 Reduced-order network models . . . 496
22.4 Digital twins and model updating . . . 501
22.5 Discrepancy and calibration . . . 504
22.6 Fidelity matched to the decision . . . 508
22.7 What to retain . . . 512
22.8 Exercises . . . 512
22.9 Connection to the TERRA framework . . . 513

23 Data and public datasets **515**
23.1 Public data and physical inference 515
23.2 Workload, power and thermal datasets 517
23.3 The ideal dataset and data quality 521
23.4 Capability as a function of instrumentation 525
23.5 Ground truth and validation 529
23.6 Question-first use of datasets 531
23.7 What to retain 535
23.8 Exercises 536
23.9 Connection to the TERRA framework 537

24 TERRA for data centers **539**
24.1 Existing tools and the inference gap 539
24.2 Use cases in diagnosis and measurement 543
24.3 Capacity, risk and flexibility 546
24.4 Boundaries of the approach 550
24.5 Outputs for operators 553
24.6 An adoption path 556
24.7 What to retain 560
24.8 Exercises 560
24.9 The book in one page 562

Notation and vocabulary

Every symbol below is used with the same meaning in every chapter.

Symbol	Unit	Meaning
P_{IT}	W	Electrical power consumed by IT equipment (servers, storage, network).
P_{facility}	W	Total electrical power drawn by the facility.
P_{cool}	W	Electrical power consumed by the cooling system.
P_{rack}, P_{server}	W	Electrical power of one rack, of one server.
P_{idle}, P^{max}	W	Power of a server at no load and at full load.
u	–	Utilization of a processor or server, between 0 and 1; also the control input of a controller, by context.
$\dot{Q}$	W	Heat flow rate.
$\dot{Q}_{\mathrm{IT}}$, $\dot{Q}_{\mathrm{rack}}$	W	Heat generated by the IT equipment, by one rack.
$\dot{Q}_{\mathrm{avail}}$, $\dot{Q}_{\mathrm{req}}$	W	Cooling capacity available, cooling required.
$\dot{Q}_{\mathrm{evap}}$, $\dot{Q}_{\mathrm{cond}}$	W	Heat absorbed at a chiller's evaporator, rejected at its condenser.
W_{comp}	W	Compressor power of a chiller.
$\dot{m}$, $\dot{m}_{\mathrm{air}}$, $\dot{m}_w$	kg/s	Mass flow rate; of air; of water.
$\dot{V}$	m^3/s	Volumetric flow rate.
c_p, $c_{p,\mathrm{air}}$, $c_{p,w}$	$\mathrm{J/(kg\,K)}$	Specific heat; of air, about $1.0\,\mathrm{kJ/(kg\,K)}$; of water, about $4.18\,\mathrm{kJ/(kg\,K)}$.
T	K or °C	Temperature.
T_{in}, T_{out}	K	Air temperature entering and leaving a server or rack.
T_{supply}, T_{return}	K	Temperature of supply and return air or water.
T_{hot}	K	Temperature of hot exhaust air.
$T_{\mathrm{CHW,s}}$, $T_{\mathrm{CHW,r}}$	K	Chilled-water supply and return temperature.
T_{amb}, T_{wb}	K	Outdoor dry-bulb and wet-bulb temperature.
T_{set}, T_{max}	K	Setpoint of a controller; upper limit of a temperature.
RH	–	Relative humidity.
Δp	Pa	Pressure difference.
K	varies	Flow-resistance coefficient in $\Delta p = K\dot{V}^2$.
α	–	Recirculation fraction: share of rack inlet air that is hot exhaust.
β	–	Bypass fraction: share of supply air that returns without passing through IT equipment.
R_{th}	K/W	Thermal resistance.
C_{th}, C	J/K	Thermal capacitance.
τ	s	Time constant, $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$.
UA	W/K	Overall conductance of a heat exchanger.
COP	–	Coefficient of performance of a chiller, $\dot{Q}_{\mathrm{evap}}/W_{\mathrm{comp}}$.

Symbol	Unit	Meaning
PUE	–	Power usage effectiveness, $P_{\text{facility}}/P_{\text{IT}}$.
N_{fan}, N_{pump}	rev/min	Rotational speed of a fan, of a pump.
η	–	Efficiency of a converter, pump or power path.
SOC, SOH	–	State of charge and state of health of a battery.
N, $N{+}1$, $2N$	–	Redundancy levels: the number of units required, one spare, two full independent sets.
x, u, d, θ	varies	State, control input, disturbance and parameter vectors of a dynamic model.
y, ϵ, b	varies	Measurement, its random error and its bias.
r	varies	Residual: measured value minus the value a model predicts.
H_i	–	A hypothesis about the cause of an observation.
H_0, H_M	–	The hypotheses of normal operation and of an inadequate model.
$\Pr(H_i \mid y)$	–	Posterior probability of a hypothesis given the measurements.
$p(x \mid y)$	varies	Posterior density of a state (or parameter) given the measurements.
$\hat{x}$, x_{MAP}	varies	An estimate of the state; the maximum a posteriori estimate.
w_k, v_k, K_k	varies	Process noise, measurement noise and Kalman gain at step k.
G	–	Topology: which component serves or feeds which.
$\mathcal{M}_j$	–	A candidate model structure.
$\mathcal{I}$	varies	Fisher information matrix of the parameters.
R_i, A	–	Reliability of a component over an interval; availability.
MTBF, MTTR	h	Mean time between failures; mean time to repair.
C, L, M	W	In Chapters 12 and 17 to 20: capacity, load and capacity margin, $M = C - L$.
C_{50}, C_{95}, C_{99}	W	Capacity exceeded with probability 50, 95 and 99 %.
c	–	A contingency: a specified failure or outage.
ϵ	–	In Chapters 12 to 20 also a risk tolerance, the largest accepted probability of a violation.
L_i, ℓ	varies	Loss or consequence of a hypothesis, scenario or action (ℓ where L is a load).
$\mathcal{R}$	varies	Expected risk: probability times consequence, summed over events.
S, a	–	A scenario; an action.
$\mathcal{H}$	bit	Entropy of a distribution over hypotheses.
IG, VOI	bit, cost	Expected information gain; expected value of information of a measurement.
Φ	–	Standard normal distribution function.
J	varies	Objective of an optimization (in Chapters 15 and 19, a sensitivity matrix).
t_{limit}	s	Time until a temperature reaches its limit at the present rate of rise.
$F^{\downarrow}$, $F^{\uparrow}$	W	Downward and upward flexibility: how far the net grid power can be lowered or raised from its baseline.

Symbol	Unit	Meaning
F_{95}	W	Firm flexibility: the flexibility available with 95 % probability.
$\Delta P(t)$	W	Flexibility trajectory: grid power minus its baseline during and after an event.
$\mathrm{CI}(t)$	kg/Wh	Carbon intensity of grid electricity.
$\mathbf{u}$, p, ρ, μ, k	varies	In Chapter 22: velocity field, pressure, density, viscosity and thermal conductivity of the air.
δ	varies	Model discrepancy: the systematic difference between a model and the system it represents.

Vocabulary

IT equipment Servers, storage and network equipment: the equipment that performs computation and whose power a data center exists to supply.

UPS Uninterruptible power supply: a converter with stored energy that keeps the IT load supplied when the upstream source fails.

PDU Power distribution unit: equipment that distributes and meters power to racks.

CRAH, CRAC Computer room air handler and computer room air conditioner: the units that cool the room air, with a chilled-water coil or with their own refrigeration cycle.

BMS, DCIM Building management system and data center infrastructure management: the supervisory software of the facility plant and of the data-center-specific infrastructure.

Measured, calculated, inferred The three kinds of statement this book keeps apart: a sensor reading, a quantity computed from readings by a known relation, and a conclusion about an unobserved state or cause.

Chapter 1

Data center system architecture: the flow of power, heat, information and control

1.1 The facility as an energy-conversion system

A data center is usually described by what it hosts: servers, storage, network equipment. This book describes it by what it does to energy. Electrical energy enters from the grid, passes through a chain of electrical equipment, is consumed by the IT equipment, and leaves the building as heat carried by air, water and finally the outdoor atmosphere. The question of this chapter is

where energy enters a data center, where it goes, what becomes heat, how that heat is removed, and what keeps the whole chain operating when a component fails.

The answer is a map of the whole facility, drawn at the level of detail that the later chapters fill in one subsystem at a time: the power path in Chapters 3 and 4, the IT load in Chapter 2, heat and air in Chapters 5 and 6, the cooling plant in Chapter 7, its dynamics in Chapter 8, sensors and controls in Chapters 9 and 10, and the efficiency metric in Chapter 11.

Jargon: IT load, UPS, PDU, CRAH, CRAC, chiller, cooling tower, BMS, DCIM, N+1

The *IT load* is the electrical power consumed by the information technology equipment, the servers, storage and network devices that perform the computation; it is the power the facility exists to supply, P_{IT}. A *UPS* (uninterruptible power supply) is a power converter with stored energy, usually batteries, that keeps the IT load supplied without interruption when the upstream source fails, for long enough for a generator to start or for the load to shut down in order. A *PDU* (power distribution unit) divides a feeder into the circuits that supply individual racks, usually with breakers and meters. A *CRAH* (computer room air handler) is a fan unit that cools room air by passing it over a coil of chilled water supplied from a central plant; a *CRAC* (computer room air conditioner) does the same job with its own refrigeration cycle and compressor inside the unit. A *chiller* is a machine that cools a water loop by moving heat out of it with a refrigeration cycle driven by an electric compressor. A *cooling tower* rejects heat from a water loop to the outdoor air, mostly by evaporating a small part of the water. The *BMS* (building management system) is the supervisory control and monitoring system of the mechanical and electrical plant; *DCIM* (data center infrastructure management) is the software that inventories, meters and monitors the data-center-specific infrastructure (racks, PDUs, power and environmental sensors). *N+1* is a redundancy level: if N units are needed to carry the load, $N + 1$ are installed, so that any one can fail or be taken out for maintenance without loss of capacity.

Key idea

A data center converts electrical energy into computation and, from the physical plant's point of view, essentially all of that electrical energy into heat. The facility is a coupled network that delivers power to the IT equipment, transports the resulting heat to the outdoor environment, watches and controls both flows, and duplicates the parts whose failure would stop them.

This section describes the three layers in which the facility is organized and states the energy balance on which the rest of the chapter rests.

1.1.1 Three layers of the facility

The facility can be drawn in three layers. The first is the *energy chain*: grid, electrical infrastructure, IT equipment, heat, cooling system, environment. Every joule that enters the building follows this chain, and its two ends are both outside the building: the energy arrives as electricity from the grid and leaves as heat to the atmosphere. The second layer *watches and controls* the first: sensors measure temperatures, powers, flows, pressures and humidity; controllers in the BMS and the DCIM compare those measurements with targets; and actuators (fans, pumps, valves, chiller staging) change the flows. The third layer gives *resilience*: uninterruptible power supplies and their batteries, standby generators, and redundant equipment of every kind, installed so that the first layer keeps working when a component of it fails.

Figure 1.1 draws the three layers. The middle row of boxes is the energy chain, read from left to right, with solid arrows for the flow of energy: electrical up to the IT equipment, thermal after it. The top row is the resilience layer; its arrows point to the parts of the energy chain each element protects. The bottom row is the sensing and control layer: a dashed path carries measurements from the energy chain to the sensors, the controls act on them, and a second dashed path carries commands from the actuators back to the cooling system. The figure shows each layer at one place for clarity; in a real facility sensors and actuators are distributed along the whole chain, and redundancy appears at almost every stage.

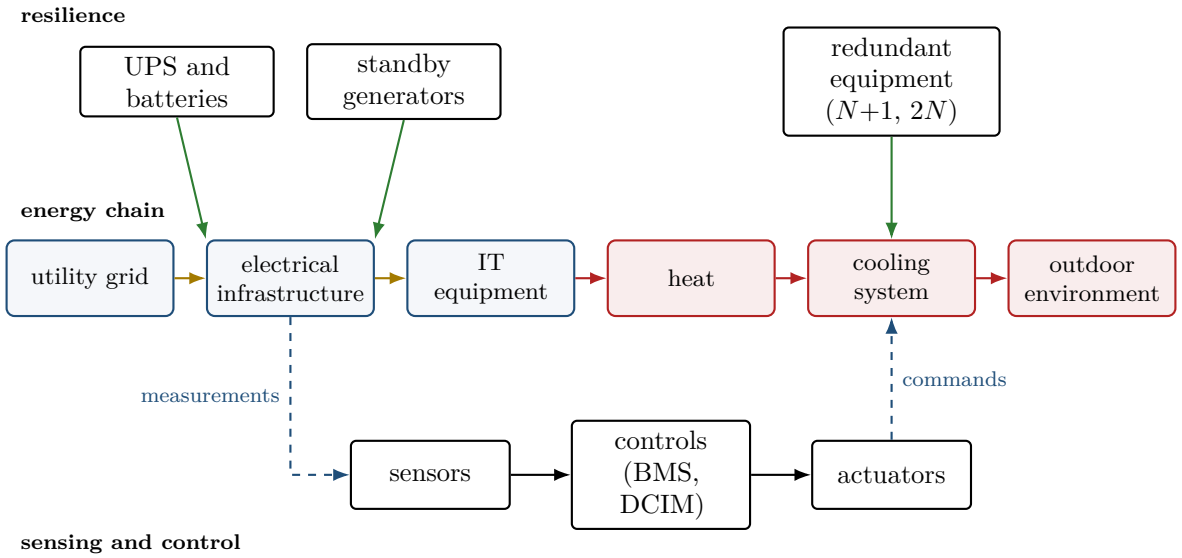


Figure 1.1: The three layers of a data center. Middle: the energy chain, electrical (gold arrows) up to the IT equipment and thermal (red arrows) after it. Top: the resilience layer and the parts of the chain it protects. Bottom: the sensing and control layer, drawn at one place for clarity although sensors and actuators are spread along the whole chain.

1.1.2 Electricity into computation and heat

What a data center does, physically, is convert electrical energy into computation and heat. The computation is the useful product, but it carries almost no energy away with it: the result of a calculation is stored as the states of transistors or sent out as signals whose power is minute compared with the power drawn to produce them. Almost all the electrical energy consumed by the IT equipment therefore ends as heat inside the building. For a server drawing P_{server},

$$\dot{Q}_{\text{server}} \approx P_{\text{server}}, \qquad \dot{Q}_{\text{IT}} \approx P_{\text{IT}}, \tag{1.1}$$

where $\dot{Q}_{\text{server}}$ is the heat flow the server releases and $\dot{Q}_{\text{IT}}$ the heat flow of all the IT equipment. A server at $P_{\text{server}} = 1\,\text{kW}$ releases about 1 kW of heat; the small exceptions, such as the optical power of signals leaving on fiber, the energy of the electrical signals on network cables and the sound of the fans, are fractions of a percent and are ignored. A room whose IT load is $P_{\text{IT}} = 1\,\text{MW}$ therefore produces about 1 MW of heat continuously, and over one day releases $1\,\text{MW} \times 24\,\text{h} = 24\,\text{MWh}$ of heat that must leave the building.

That statement fixes the job of the cooling system. Its job is not to "make the room cold". The room is at a steady temperature only when heat leaves it as fast as it is generated, and the cooling system is the set of equipment that *transports* heat from the IT equipment to the outdoor environment at the rate the IT equipment produces it. A cooling system that removed 0.95 MW from a 1 MW room would not keep it slightly warmer; it would let the room's temperature rise without limit, at a rate set by how much thermal mass the room contains (Section 1.6.4 estimates that rate). Thinking of cooling as transport, rather than as the production of cold, is what makes the rest of the facility readable as a set of conservation laws.

Principle 1.1. *Essentially all the electrical power consumed by IT equipment becomes heat, $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$. The cooling system does not produce cold; it transports that heat to the outdoor environment, and in steady operation it must carry it at the rate it is produced.*

1.2 The electrical and IT subsystems

Before it becomes heat, the energy passes through two subsystems: the electrical infrastructure that brings it from the grid to the servers, and the IT equipment that consumes it. This section follows the power path, states its two jobs, and describes how IT equipment is grouped from a server up to a data hall.

1.2.1 The power path from grid to server

Electrical power reaches a server through a chain of equipment:

$$\text{utility grid} \rightarrow \text{switchgear} \rightarrow \text{transformer} \rightarrow \text{UPS} \rightarrow \text{PDU} \rightarrow \text{rack} \rightarrow \text{server}.$$

The utility delivers power at medium voltage, typically between about 10 and 35 kV, depending on the country and the size of the site. *Switchgear* is the assembly of breakers, switches and protection relays that connects, disconnects and protects the incoming feeders and the circuits downstream. A *transformer* steps the voltage down to the low voltage used inside the building, typically 400 or 480 V three-phase. The UPS conditions the power and holds it up through interruptions, the PDU divides it among racks, and inside each rack power strips feed the individual servers, whose power supplies convert it to the low direct-current voltages, typically 12 V and below, that processors, memory and drives use. The voltages quoted are typical values; the arrangement of the chain is the same everywhere.

Figure 1.2 draws the chain. The top row of boxes is the path of power, left to right, from the utility to the server. The boxes below it are the elements that keep the path working when part of it fails: a second utility feeder and standby generators connected through transfer equipment at the switchgear, and batteries connected to the UPS. The small red arrows above the transformer, the UPS and the PDU mark the conversion losses of each stage, which leave as heat in the electrical rooms rather than in the data hall. Every stage of the path is a converter or a switch, and each stage is a place where power is measured, where it is lost, and where a failure can interrupt the flow.

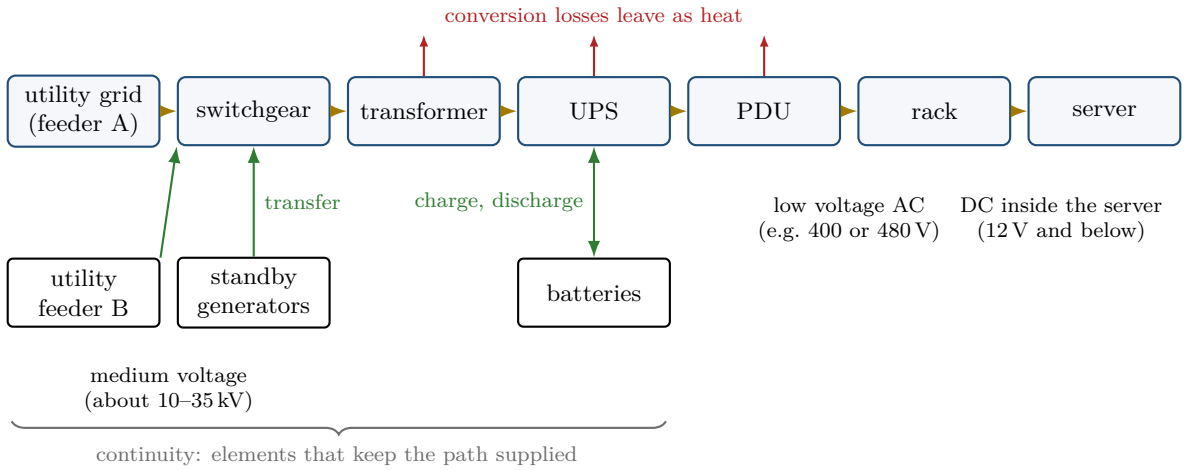


Figure 1.2: The electrical power path from the utility to a server (top row, gold arrows) and the equipment that keeps it supplied when part of it fails (bottom row, green arrows). Red arrows mark the conversion losses of the transformer, the UPS and the PDU. Voltages are typical values.

1.2.2 The two jobs of the electrical system

The electrical system has two jobs. The first is *delivery*: bring power at the right voltage and quality to every server, at whatever rate the servers draw it. The second is *continuity*: keep delivering when something upstream fails, whether that is the utility supply, a transformer, a breaker or a cable. A server does not tolerate even a short interruption: its power supply holds up its output for only a few tens of milliseconds, so an interruption longer than that restarts the server and loses the work in progress.

Continuity is the reason for most of the equipment in Figure 1.2 that is not on the direct path. The UPS and its batteries carry the load through the first seconds and minutes of an outage; batteries are commonly sized for a few minutes at full load. Standby generators, usually diesel, start when the utility fails and take over the load once they have reached speed, typically within ten to fifteen seconds. Redundant utility feeders, breakers that isolate a fault so that the rest of the system keeps running, and transfer equipment that switches the load from one source to another complete the arrangement. Chapter 3 follows the distribution path in detail and Chapter 4 the backup equipment; at the level of this chapter it is enough that the electrical subsystem is both a delivery network and a set of parallel paths kept ready for failure.

1.2.3 From server to data hall

The IT equipment is organized in a hierarchy:

$$\text{server} \rightarrow \text{rack} \rightarrow \text{cluster} \rightarrow \text{data hall}.$$

A *rack* is a standard cabinet, about 0.6 m wide and 1.0 to 1.2 m deep, holding a few to several dozen servers together with their network switches and power strips. A *cluster* is a group of racks that work together on the same workloads, and a *data hall* is the room, or a partitioned zone of a room, that contains many racks together with their cooling units. A rack consumes an electrical power P_{rack} and, by equation (1.1), releases a heat flow

$$\dot{Q}_{\text{rack}} \approx P_{\text{rack}}. \tag{1.2}$$

A rack at $P_{\text{rack}} = 40\,\text{kW}$ releases roughly 40 kW of heat that must be removed from the space it occupies.

The power of a rack matters not only through the total but through its concentration. Conventional enterprise racks have long drawn about 5 to 10 kW each; racks of accelerators for artificial-intelligence workloads draw 40 kW and more, and the densest designs exceed 100 kW. A 40 kW rack with a footprint of about $0.6 \times 1.2 = 0.72\,\text{m}^2$ releases about $40/0.72 \approx 56\,\text{kW}$ per square meter of rack footprint, the heat of a few dozen domestic electric heaters in a single cabinet. High density does not change the energy balance of the facility, but it concentrates the heat spatially, and the air or liquid that carries it away must be delivered to that one place at a correspondingly high rate. Section 1.3.1 quantifies the rate.

1.3 The path of heat

Once the electrical energy has become heat in the IT equipment, it travels through a sequence of media and machines to the outdoor environment. This section follows that path stage by stage: air, the room's airflow arrangement, the cooling unit, the chilled-water loop, the chiller, the condenser loop and the cooling tower.

1.3.1 Air as the first heat-transport medium

In a conventional air-cooled rack, cool air enters at the front of each server, the server's fans pull it across the processors, accelerators, memory and power supplies, and the air leaves at the

rear, warmer by the heat it has picked up. The air is the first medium that carries the heat away from the equipment. The heat it carries follows from the energy balance of the air stream: each kilogram of air that warms by one kelvin absorbs $c_{p,\mathrm{air}}$ joules, so a mass flow $\dot{m}_{\mathrm{air}}$ warming from T_{in} to T_{out} carries

$$\dot{Q} = \dot{m}_{\mathrm{air}}\, c_{p,\mathrm{air}}\, (T_{\mathrm{out}} - T_{\mathrm{in}}), \tag{1.3}$$

where $c_{p,\mathrm{air}} \approx 1.0\,\mathrm{kJ/(kg\,K)}$ is the specific heat of air. For $\dot{m}_{\mathrm{air}} = 2\,\mathrm{kg/s}$ and a temperature rise of $10\,\mathrm{K}$,

$$\dot{Q} = 2\,\mathrm{kg/s} \times 1.0\,\mathrm{kJ/(kg\,K)} \times 10\,\mathrm{K} = 20\,\mathrm{kW}.$$

With air at a density of about $1.2\,\mathrm{kg/m^3}$, $2\,\mathrm{kg/s}$ is about $1.7\,\mathrm{m^3/s}$ of air.

Read the other way, equation (1.3) states the airflow that a given heat requires. The $40\,\mathrm{kW}$ rack of Section 1.2.3, with the same $10\,\mathrm{K}$ rise, needs $\dot{m}_{\mathrm{air}} = 40/(1.0 \times 10) = 4\,\mathrm{kg/s}$, about $3.3\,\mathrm{m^3/s}$, through one cabinet; at $2\,\mathrm{kg/s}$ its air would have to warm by $20\,\mathrm{K}$ instead. The rise is limited by the temperatures the components tolerate and the air can be moved only so fast through a cabinet, which is why high-density racks press against the limits of air as a transport medium. Chapter 5 develops this balance and Chapter 6 the movement of the air.

1.3.2 Hot aisles and cold aisles

If racks were placed in random orientations, the hot exhaust of one rack would blow into the intake of the next. The intake air would be a mixture of cool supply and hot exhaust, the servers would run hotter for the same supply temperature, and the cooling units would receive return air that is cooler than the exhaust and therefore carries less heat per kilogram. This is inefficient, because the cooling system must move more air and supply it colder to achieve the same intake temperature, and potentially dangerous, because a server that breathes another server's exhaust can exceed its temperature limit. To prevent it, racks are arranged in rows with their fronts facing each other across a *cold aisle* and their backs facing each other across a *hot aisle*. The air then follows a loop,

$$\text{cooling unit} \rightarrow \text{cold aisle} \rightarrow \text{servers} \rightarrow \text{hot aisle} \rightarrow \text{cooling unit},$$

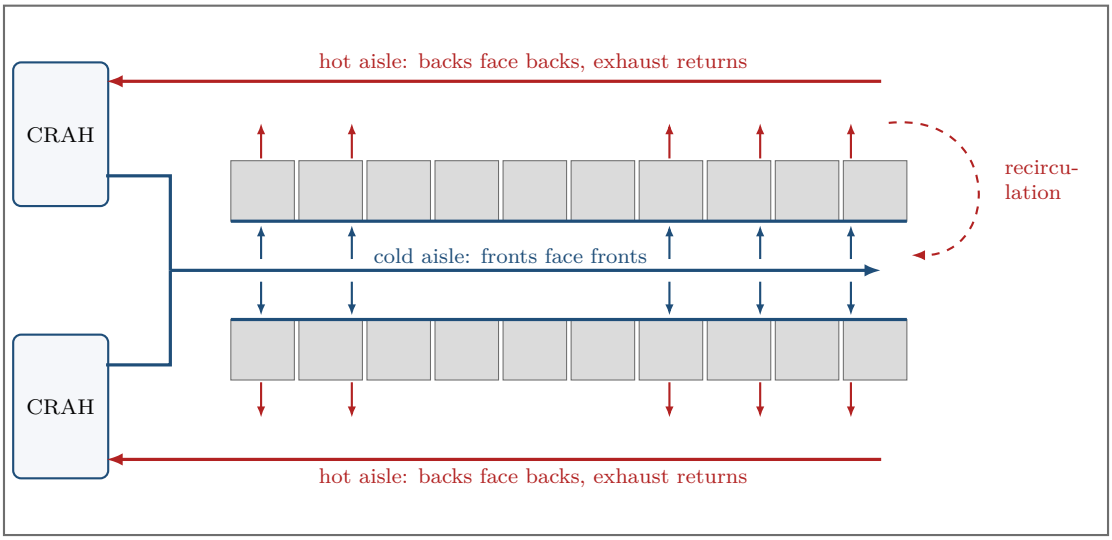
and the arrangement, standard in air-cooled facilities, reduces the mixing of supply and exhaust.

Figure 1.3 shows a plan view of such a room, seen from above. Two rows of racks face each other across the cold aisle in the middle; the hot aisles lie behind them, at the top and the bottom of the plan. Cooling units on the left wall push cool supply air (blue) into the cold aisle, the servers draw it through the racks, and the warm exhaust (red) collects in the hot aisles and returns to the cooling units. The dashed red path at the right end of the upper row is *recirculation*: exhaust that finds its way around the end of the row, or over the top of the racks, back into the cold aisle. The reverse leak, supply air that returns to the cooling units without passing through any server, is *bypass*. Chapter 6 quantifies both through the recirculation fraction α and the bypass fraction β. For an illustration of the cost of mixing: if $\alpha = 0.2$ of the air entering a server is exhaust at $37^{\circ}\mathrm{C}$ and the rest is supply at $22^{\circ}\mathrm{C}$, the server's intake is $0.8 \times 22 + 0.2 \times 37 = 25^{\circ}\mathrm{C}$, three kelvin warmer than the supply.

Air as an intermediary. The arrangement makes explicit a point that is easy to miss: the building's cooling system does not cool the server directly. It cools the air, and the air cools the server,

$$\text{server} \rightarrow \text{air} \rightarrow \text{cooling equipment}.$$

Air is a heat-transport medium between the two, and every property of the air path (how much air each rack receives, how much of it is mixed with exhaust, how much bypasses the servers) sits between the cooling equipment and the temperature the server experiences.



plan view; thick blue edge = rack fronts (air intakes); blue = cool supply air, red = warm exhaust

Figure 1.3: Hot-aisle and cold-aisle arrangement in plan view. Cooling units (left) supply cool air to the cold aisle between two rows of racks whose fronts face each other; servers draw it through and exhaust into the hot aisles behind the rows, from which it returns to the cooling units. The dashed path is recirculation of exhaust around the row end into the cold aisle.

1.3.3 The computer room air handler and air conditioner

The cooling unit at the end of the air loop takes warm return air from the hot aisle and delivers cool supply air to the cold aisle:

$$\text{warm return air} \rightarrow \text{CRAH or CRAC} \rightarrow \text{cool supply air.}$$

Inside a chilled-water CRAH, fans push the return air across a heat exchanger, a coil of finned tubes through which cold water flows. The air gives up heat to the metal of the coil and the metal gives it to the water, so the heat moves from the air to the chilled water and the air leaves the coil cooler. A CRAC achieves the same result with its own refrigeration cycle: the cold surface is the evaporator of a compressor circuit inside the unit, and the heat is rejected through a separate condenser, often outdoors.

With a chilled-water CRAH, the path of the heat has so far been

$$\text{server} \rightarrow \text{air} \rightarrow \text{water.}$$

The CRAH is the first point at which the heat changes medium inside the building. The air side of the unit carries the room's heat at low density and over short distances; the water side carries it in pipes over the long distance to the central plant. The fans of the CRAH also consume electricity, and that electricity becomes heat in the air stream that the coil must remove as well, a first example of the cooling system adding to the heat it handles.

1.3.4 The chilled-water loop

The water that leaves the CRAH coils circulates in a closed loop, the *chilled-water loop*, between the coils and the chillers. In a typical design the coil receives water at a supply temperature $T_{\mathrm{CHW,s}} = 7^{\circ}\mathrm{C}$ and returns it at $T_{\mathrm{CHW,r}} = 12^{\circ}\mathrm{C}$. The heat it carries follows from the same balance as for air, written for water,

$$\dot{Q} = \dot{m}_w \, c_{p,w} \, (T_{\mathrm{return}} - T_{\mathrm{supply}}) = \dot{m}_w \, c_{p,w} \, (T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}}), \tag{1.4}$$

with $c_{p,w} \approx 4.18\,\mathrm{kJ/(kg\,K)}$ the specific heat of water. To carry the 1 MW of a 1 MW IT load with a 5 K rise, the loop needs

$$\dot{m}_w = \frac{1000\,\mathrm{kW}}{4.18\,\mathrm{kJ/(kg\,K)} \times 5\,\mathrm{K}} \approx 47.8\,\mathrm{kg/s},$$

about 48 liters per second. The same heat carried by air with a 10 K rise needs $\dot{m}_{\mathrm{air}} = 1000/(1.0 \times 10) = 100\,\mathrm{kg/s}$, about $83\,\mathrm{m^3/s}$. Per unit volume and per kelvin, water carries about $4180\,\mathrm{kJ/(m^3\,K)}$ and air about $1.2\,\mathrm{kJ/(m^3\,K)}$, a ratio of about 3500; that ratio is why heat is moved over long distances by water and why the densest racks bring liquid close to the chips.

At the rack the heat flow is written $\dot{Q} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}}\,\Delta T_{\mathrm{air}}$; at the loop it is written $\dot{Q} = \dot{m}_w c_{p,w}\,\Delta T_w$. They are two measurements of one quantity. Aside from losses, the heat added by fans and pumps, small gains through the building envelope, and heat stored while temperatures change, the heat released by the IT equipment, the heat carried by the air, and the heat carried by the chilled water are the same flow:

$$\dot{Q}_{\mathrm{IT}} \approx \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}}\,(T_{\mathrm{out}} - T_{\mathrm{in}}) \approx \dot{m}_w c_{p,w}\,(T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}}). \tag{1.5}$$

Equation (1.5) is the beginning of a physics-constrained model of the facility. Each of its three members can be computed from different meters: electrical power, air temperatures and flows, water temperatures and flows. Because they must agree, a disagreement among them is information, either about a meter that is wrong or about a heat path that has not been accounted for.

Principle 1.2. *The heat released by the IT equipment, the heat carried by the air and the heat carried by the chilled water are the same flow, up to losses, added fan and pump heat, and storage. Computed from independent meters, they must agree, and their disagreement is a measurement.*

1.3.5 The chiller

The chilled-water loop delivers warm water to the chiller and receives cold water back:

$$\text{warm chilled water } (12^\circ\mathrm{C}) \rightarrow \text{chiller} \rightarrow \text{cold chilled water } (7^\circ\mathrm{C}).$$

The chiller does not destroy heat; no machine does. It moves heat from the chilled-water loop to a second, warmer water loop, the condenser loop, and since heat does not flow by itself from a colder body to a warmer one, moving it in that direction requires work. The chiller is a vapor-compression refrigeration machine: a refrigerant evaporates at low pressure in a heat exchanger called the *evaporator*, absorbing the heat $\dot{Q}_{\mathrm{evap}}$ from the chilled water; an electric compressor raises the refrigerant's pressure, and with it its condensing temperature, using the power W_{comp}; and the refrigerant condenses at high pressure in a second heat exchanger, the *condenser*, releasing the heat $\dot{Q}_{\mathrm{cond}}$ into the condenser water. Energy conservation for the refrigerant loop in steady operation gives

$$\dot{Q}_{\mathrm{cond}} = \dot{Q}_{\mathrm{evap}} + W_{\mathrm{comp}}, \qquad \mathrm{COP} = \frac{\dot{Q}_{\mathrm{evap}}}{W_{\mathrm{comp}}}, \tag{1.6}$$

where COP, the coefficient of performance, is the heat moved per unit of compressor work. The heat to be rejected outdoors is the IT heat plus the chiller's own work.

With $\dot{Q}_{\mathrm{evap}} = \dot{Q}_{\mathrm{IT}} = 1\,\mathrm{MW}$ and $W_{\mathrm{comp}} = 200\,\mathrm{kW}$, the chiller has $\mathrm{COP} = 1000/200 = 5$, and the condenser must reject

$$\dot{Q}_{\mathrm{cond}} = 1.0\,\mathrm{MW} + 0.2\,\mathrm{MW} = 1.2\,\mathrm{MW},$$

more heat than the servers generated. The compressor's electricity is one of the facility's largest loads after the IT equipment itself, and it enters the building's heat budget twice: once on the electricity bill and once as heat that the rest of the plant must carry outdoors. Chapter 7 describes the chiller and the factors that set its COP.

1.3.6 The cooling tower and the hierarchy of loops

In a water-cooled plant the condenser water carries the 1.2 MW from the chillers to cooling towers outdoors. In an illustrative design it leaves the chiller condensers at 35°C and returns from the towers at 30°C, so by equation (1.4) its flow is $1200/(4.18 \times 5) \approx 57.4\,\mathrm{kg/s}$, larger than the chilled-water flow because it carries the compressor work as well. In the tower the warm water is sprayed over a fill material through which fans draw outdoor air, and a small part of the water evaporates. Evaporation absorbs about 2400 kJ per kilogram of water, so rejecting 1.2 MW mostly by evaporation consumes about $1200/2400 = 0.5\,\mathrm{kg/s}$ of water, roughly $43\,\mathrm{m}^3$ per day, which is replaced as make-up water. Because the tower cools by evaporation, the lowest temperature it can reach is set by the outdoor wet-bulb temperature T_{wb}, not by the dry-bulb temperature T_{amb}. Plants without towers reject the condenser heat through air-cooled condensers or dry coolers instead, at the price of higher condensing temperatures on hot days.

The complete path of the heat in a water-cooled plant is a chain of loops,

$$\text{server} \to \text{air} \to \text{chilled water} \to \text{chiller} \to \text{condenser water} \to \text{cooling tower} \to \text{atmosphere},$$

a hierarchy of heat-transfer loops, each of which picks up heat from the loop before it and hands it to the loop after it, until the last one rejects it outdoors. Each loop has its own fluid, its own flow, its own pumps or fans, and its own temperature range.

Figure 1.4 draws the chain. The boxes are the stages the heat passes through, read along the arrows: left to right along the top row from the chip to the chilled-water loop, then right to left along the bottom row from the chiller to the atmosphere. Under each box is its illustrative temperature level. The brace above the top row marks the stages that carry the IT heat, 1.0 MW; the box at the lower right is the compressor power that joins the flow in the chiller; and the brace below the bottom row marks the stages that carry 1.2 MW. The figure shows two facts at once: the heat flow is the same at every stage until the chiller adds its work, and the temperature falls from the chip to the chilled water and then is raised by the chiller to a level above the outdoor wet-bulb temperature, so that it can flow outdoors.

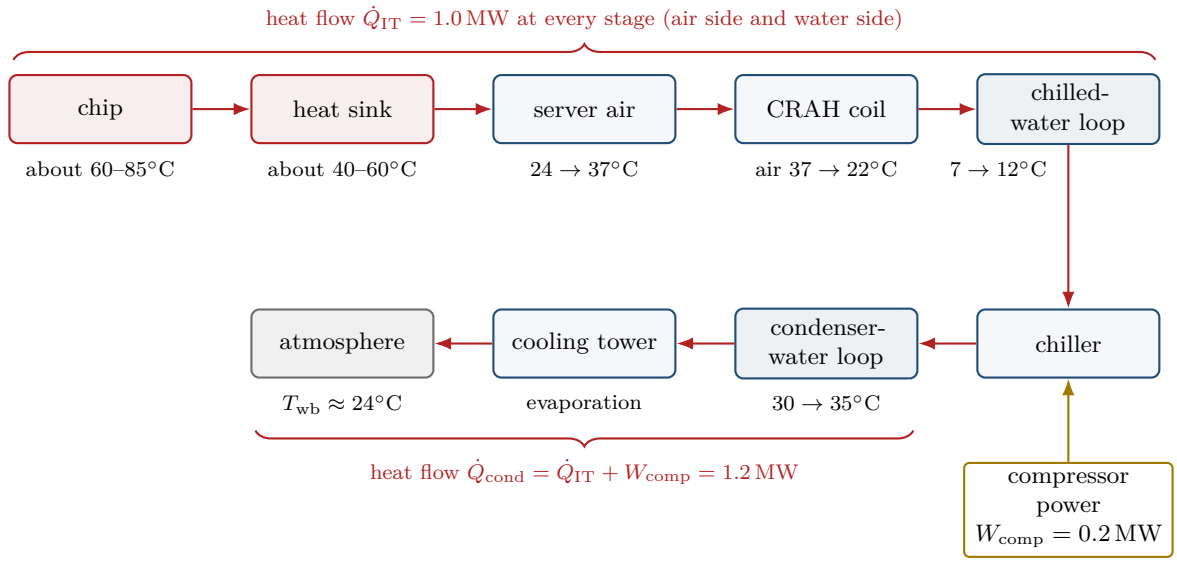


Figure 1.4: The heat-rejection chain of a water-cooled data center, from chip to atmosphere, with illustrative temperature levels under each stage. The IT heat, 1.0 MW, is the same at every stage up to the chiller; the chiller adds its compressor work, 0.2 MW, and the condenser loop, tower and atmosphere carry 1.2 MW (equation (1.6)).

1.3.7 Following one joule

Following a single joule through the facility summarizes the section. One joule of electricity enters a server and is consumed in its transistors, becoming about one joule of heat. That joule moves by conduction from the chip into its heat sink, by convection from the heat sink into the air the server's fans draw through, with the air into the hot aisle and the CRAH, across the CRAH coil into the chilled water, with the chilled water to the chiller's evaporator, through the refrigerant to the condenser, where it is joined by the compressor's work, into the condenser water, and with it to the cooling tower, where it leaves in evaporated water and warmed air to the atmosphere:

chip → heat sink → air → CRAH coil → chilled water → chiller → condenser loop → outside.

Every step of that path except one moves heat from a warmer place to a colder one, which happens by itself provided there is a path and a flow. The one exception is the chiller, which lifts the heat from about 7–12°C to 30–35°C so that it can flow to an atmosphere at a wet-bulb temperature of 24°C. The data center is therefore an energy-transport network: a set of paths along which a conserved quantity, energy, flows from a source to a sink, driven by temperature differences and moved by fans, pumps and one compressor. Each stage has a capacity, the largest flow it can carry at the temperatures available to it, and the network as a whole can carry no more heat than its most constrained stage.

1.4 Three coupled networks and the control loop

The energy chain of the previous sections is one network. A data center contains three, and they are coupled to each other. This section names the three, describes the feedback loop by which control keeps the thermal network within limits, and traces how a change in workload propagates through all of them.

1.4.1 Electrical, thermal-fluid and information networks

The *electrical network* runs from the grid through switchgear, UPS and PDU to the racks. Its quantity is power P, in kilowatts, and its laws are those of circuits: current divides at junctions and power is conserved up to conversion losses. The *thermal-fluid network* runs from the racks through the air to the water and the environment. Its quantities are temperatures T, mass flows $\dot{m}$, heat flows $\dot{Q}$ and pressures p, and its laws are conservation of mass for the air and water, conservation of energy for the heat, and the pressure-flow relations of ducts, pipes, fans and pumps. The *information and control network* runs from the sensors to the controllers to the actuators. Sensors report temperatures, electrical powers, flows, relative humidity RH, pressure differences Δp and equipment states (running, stopped, alarmed); controllers compute commands; actuators manipulate fan speed, pump speed, valve position and chiller staging, the number of chillers running. Its quantities are signals, not energy, and its connections are links between devices rather than pipes or cables carrying power.

Figure 1.5 draws the three networks as stacked bands. The top band is the electrical network, read from left to right; the middle band is the thermal-fluid network, read from right to left, from the rack to the environment; the bottom band is the information and control network. The vertical solid arrow on the right is the main coupling between the top two: the IT power becomes the IT heat. The dashed gold arrows are the second coupling: the electrical network also supplies the cooling equipment. The dashed blue arrows connect the bottom band to the others: measurements flow down into the sensors, commands flow up from the actuators to the fans, pumps, valves and chillers. The three networks share nodes, a rack is a node of all three, and none can be analyzed alone: the plant, the sensors, the controllers and the actuators form one loop,

physical plant → sensors → controllers → actuators → physical plant.

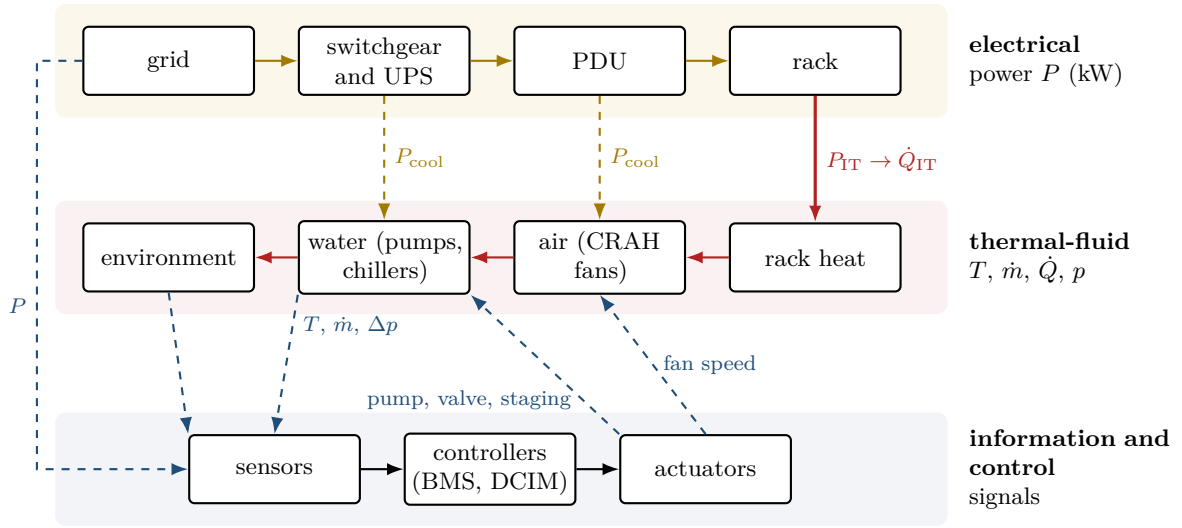


Figure 1.5: The three coupled networks of a data center. The electrical network (top) delivers power; the thermal-fluid network (middle) carries heat from the racks to the environment; the information and control network (bottom) measures both and acts on the cooling equipment. The solid red arrow is the conversion of IT power into heat; the dashed gold arrows are the electrical supply of the cooling equipment.

1.4.2 The control loop

The networks are coupled most visibly through control. Suppose a rack's inlet temperature rises to 29°C, above an illustrative setpoint of 27°C, the upper end of the inlet range commonly recommended for air-cooled IT equipment. The inlet has risen because the cold aisle in front of the rack is receiving less cool air than the servers draw, so part of their intake is recirculated exhaust. A controller in the BMS compares the measurement with the setpoint and raises the speed command u of the CRAH fans. The fans deliver more air, $\dot{m}_{\text{air}}$ rises, the cold aisle fills with supply air and displaces the recirculated exhaust, and the coil's capacity to carry heat, $\dot{Q} = \dot{m}_{\text{air}} c_{p,\text{air}} \Delta T$, rises with the flow. The inlet temperature falls toward the setpoint, the controller sees the smaller error and stops raising the fan speed. The chain

$$T \to \text{sensor} \to \text{controller} \to \text{fan} \to \dot{m}_{\text{air}} \to T$$

is a feedback loop: the output of the process, a temperature, is measured and used to adjust the input that drives it.

Figure 1.6 draws the loop as a block diagram. Signals travel along the arrows. From the left, the setpoint T_{set} enters a comparison (the circle), which subtracts the measured temperature y and passes the error e to the controller. The controller's output is the fan-speed command u; the CRAH fan turns it into an airflow $\dot{m}_{\text{air}}$; the cold aisle and rack, whose heat balance is the process being controlled, turn the airflow into an inlet temperature T_{in}. The disturbances d, drawn entering the process from above, are the inputs nobody commands: the IT load, recirculation paths and the supply-air temperature. The lower path is the feedback: the sensor measures T_{in} and returns the measurement y to the comparison. The loop acts on the error, so it corrects the effect of any disturbance without needing to know which one occurred.

The state form. The loop is one instance of a general model. Writing the facility's physical state as a vector x (temperatures of air volumes, coils, water loops and equipment masses; the

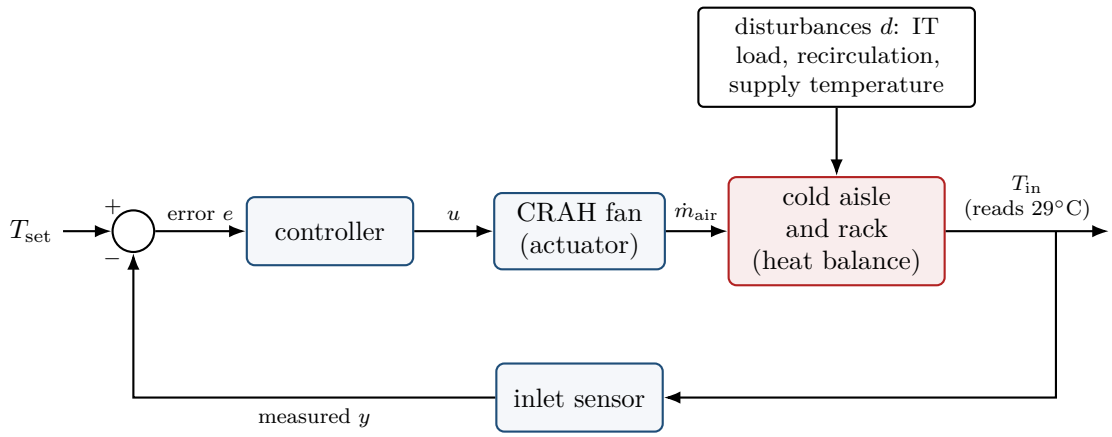


Figure 1.6: The rack-inlet temperature control loop. The controller acts on the error between the setpoint and the measured inlet temperature and commands the CRAH fan speed; the airflow changes the inlet temperature through the heat balance of the cold aisle and rack, which disturbances also drive. Values are illustrative: $T_{\text{set}} = 27°\text{C}$, measured inlet 29°C.

state of charge of the UPS batteries), its controllable inputs as u (fan speeds N_{fan}, pump speeds N_{pump}, valve positions, chiller staging), the inputs nobody controls as d (IT load, outdoor temperatures T_{amb} and T_{wb}), and the physical parameters that are not known exactly as θ (coil conductance UA, recirculation fraction α, thermal resistances, sensor biases), the facility obeys

$$\dot{x} = f(x, u, d, \theta), \qquad y = g(x, \theta) + b + \epsilon, \tag{1.7}$$

where the first equation is the physics, conservation of energy and mass written for each node, and the second says what the sensors report: a function of the state, plus a bias b and a random error ϵ. The controller sees only y. The modeler wants x and θ. The gap between them is the subject of the second half of this book.

1.4.3 Disturbances and the coupling of cooling back into load

A disturbance propagates through all three networks. When the workload rises, the IT power rises, the IT heat rises with it by equation (1.1), rack temperatures begin to rise, the cooling controls respond by raising the air or water flow, and the cooling equipment draws more electricity:

$$\text{workload} \uparrow \to P_{\text{IT}} \uparrow \to \dot{Q}_{\text{IT}} \uparrow \to T_{\text{rack}} \uparrow \to \dot{m}_{\text{air}}, \dot{m}_w \uparrow \to P_{\text{cool}} \uparrow.$$

Because cooling itself consumes electricity, the chain closes on itself:

$$\text{compute demand} \to \text{electrical load} \to \text{heat} \to \text{cooling demand} \to \text{electrical load}.$$

This is a feedback-like coupling between the electrical and thermal networks. It does not run away, because the cooling power is a fraction of the heat it removes, but it amplifies every change in IT load: the facility's electrical demand moves by more than the IT load does.

The facility's electrical demand is the sum of the IT load and everything that supports it,

$$P_{\text{facility}} = P_{\text{IT}} + P_{\text{cool}} + P_{\text{loss}} + P_{\text{other}}, \tag{1.8}$$

where P_{loss} is the conversion loss of the power path (transformers, UPS, PDUs, cables) and P_{other} is lighting and auxiliary loads. As an illustration, a facility with $P_{\text{IT}} = 1.00\,\text{MW}$ and $P_{\text{facility}} = 1.45\,\text{MW}$ might divide its 0.45 MW of overhead into 0.20 MW of chiller compressor power, 0.15 MW of fans, pumps and tower fans, 0.08 MW of electrical losses and 0.02 MW of

lighting and other loads. If the IT load rises by 10 %, to 1.10 MW, and the cooling power scales in proportion to the heat it removes, P_{cool} rises from 0.35 to about 0.385 MW; the conversion losses also grow, so the facility demand rises by more than the 0.10 MW of IT load. Fan power rises faster than in proportion when the airflow must rise, because it grows roughly with the cube of fan speed (Chapter 6), and every watt of fan and pump power ends as heat in the air or water that the plant must remove as well. The ratio $P_{\text{facility}}/P_{\text{IT}}$, here 1.45, is the facility's power usage effectiveness, the subject of Chapter 11.

1.5 Redundancy and measurement

The resilience layer and the measurement layer of Figure 1.1 each change how the facility must be described. This section treats redundancy as parallel paths and states the question it answers as a probability, then lists what a facility measures, what it does not, and how the two are related.

1.5.1 Parallel paths and redundancy levels

Reliability adds parallel paths to every network. The electrical network has UPS A and UPS B, each feeding one of the two power cords of a dual-corded server; the cooling plant has chillers 1, 2 and 3 and often a fourth; the air side has more CRAH units than the load needs. The redundancy level is named by comparison with N, the number of units required to carry the design load. A design at N installs exactly the required units; any failure reduces capacity below the load. A design at $N{+}1$ installs one spare, so any single unit can fail or be taken out for maintenance. A design at $2N$ installs two complete, independent sets, each able to carry the whole load, so an entire set can be lost, as in the A and B electrical paths.

Table 1.1 works one example. The required cooling is 3 MW and each chiller provides 1 MW, so $N = 3$. The table lists, for each design, the number of chillers installed, the installed capacity, and the capacity left after one and after two chillers fail. Only the N design falls short after one failure; $N{+}1$ survives one failure exactly at the required capacity and falls short after a second; $2N$ survives three failures. The same counting appears in contingency analysis of a transmission system, where a network is planned to carry its load after the loss of any one element.

Table 1.1: Redundancy levels for a required cooling of 3 MW provided by 1 MW chillers. Capacities below the requirement are marked as short.

Design	Chillers installed	Installed capacity	Capacity, one unit out	Capacity, two units out
N	3	3 MW	2 MW (short)	1 MW (short)
$N{+}1$	4	4 MW	3 MW	2 MW (short)
$2N$	6	6 MW	5 MW	4 MW

1.5.2 Capacity adequacy as a probability

Table 1.1 answers a deterministic question: whether the installed capacity exceeds the design load after a stated number of failures. The operational question is different. The IT load is uncertain and varies by hour; each chiller's capacity falls on hot days, when its condenser runs warmer; equipment degrades; and at any moment one chiller may already be out for maintenance. The question is then not whether capacity exceeds load, but how likely it is that it does not:

$$\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}}), \tag{1.9}$$

the probability that the available cooling capacity $\dot{Q}_{\text{avail}}$ is below the required cooling $\dot{Q}_{\text{req}}$, given the uncertainty in the IT load, the outdoor temperature, the condition of the equipment and the units unavailable.

An illustrative calculation shows how the probability differs from the count. Let each chiller of the $N{+}1$ plant be available, independently of the others, with probability 0.98. On a mild day each chiller delivers its full 1 MW, and the plant falls short only if fewer than three of the four are available:

$$\Pr(\text{short} \mid \text{mild}) = 1 - \left[0.98^4 + 4 \times 0.98^3 \times 0.02\right] = 1 - (0.9224 + 0.0753) = 0.0023.$$

On a hot day, suppose each chiller's capacity falls to 0.9 MW. Three chillers now give 2.7 MW, less than the 3 MW required, so the plant falls short unless all four are available:

$$\Pr(\text{short} \mid \text{hot}) = 1 - 0.98^4 = 0.0776.$$

If hot conditions occur in 10 % of hours, the probability of a shortfall in a randomly chosen hour is $0.9 \times 0.0023 + 0.1 \times 0.0776 \approx 0.0099$, about four times the mild-day value, and almost all of it comes from hot hours. For comparison, an N plant of three chillers on a mild day falls short with probability $1 - 0.98^3 = 0.0588$. The $N{+}1$ label is the same in both weathers; the risk is not. The numbers are illustrative, and the independence assumption is optimistic, because failures can share a cause (a common electrical supply, a heat wave), a point a later chapter on reliability develops.

Principle 1.3. *Redundancy is a count of parallel units; adequacy is a probability. The operational question is* $\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}})$ *under uncertain load, weather, degradation and unavailable units, and an* $N{+}1$ *design can carry very different risks in different conditions.*

1.5.3 What is measured and what is not

A facility's monitoring system reports many quantities, but far from all of those that matter. Table 1.2 lists an illustrative set of readings a facility might have, followed by quantities calculated from them. The first column names the quantity, the second gives its symbol, the third its value, and the fourth says whether the value is measured or calculated, and from what. The six measured values are the facility power $P_{\text{facility}} = 1.45\,\text{MW}$, the IT power $P_{\text{IT}} = 1.00\,\text{MW}$, a rack inlet temperature of 24°C, the same rack's outlet temperature of 37°C, and chilled-water supply and return temperatures of 7 and 12°C.

The list of what the facility does not know is longer. It typically does not know the airflow through every rack, the amount of heat recirculation in each aisle, the effectiveness of each cooling coil, whether every sensor is unbiased, the true cooling capacity of the plant as it stands today, the thermal resistances between components (chip to heat sink, heat sink to air), or where energy that no submeter records actually goes. These are exactly the quantities that decide whether the facility is operating safely and efficiently, and none of them is read from a single sensor.

1.5.4 Measured, calculated and inferred

Three kinds of statement can be made about the facility, and they are very different claims. A *measured* statement is a sensor reading: $T_{\text{in}} = 24°\text{C}$. It can be wrong, through bias, drift or a misplaced sensor, but it is a direct report. A *calculated* statement applies a known relation to measured values: $\dot{Q} = \dot{m}_{\text{air}} c_{p,\text{air}} (T_{\text{out}} - T_{\text{in}})$, or the 1.54 kg/s airflow of Table 1.2. Its validity depends on the relation and on the assumptions behind it; the airflow calculation assumes that

Table 1.2: An illustrative set of facility readings and the quantities calculated from them. A calculated value is only as good as the relation and the assumption it rests on.

Quantity	Symbol	Value	Kind
Facility power	P_{facility}	1.45 MW	measured (utility meter)
IT power	P_{IT}	1.00 MW	measured (UPS or PDU meters)
Rack inlet air temperature	T_{in}	24°C	measured
Rack outlet air temperature	T_{out}	37°C	measured
Chilled-water supply temperature	$T_{\text{CHW,s}}$	7°C	measured
Chilled-water return temperature	$T_{\text{CHW,r}}$	12°C	measured
Facility overhead	$P_{\text{facility}} - P_{\text{IT}}$	0.45 MW	calculated by subtraction
Power usage effectiveness	PUE	1.45	calculated, $P_{\text{facility}}/P_{\text{IT}}$
Rack air temperature rise	$T_{\text{out}} - T_{\text{in}}$	13 K	calculated by subtraction
Chilled-water temperature rise	$T_{\text{CHW,r}} - T_{\text{CHW,s}}$	5 K	calculated by subtraction
Chilled-water flow	$\dot{m}_w$	47.8 kg/s	calculated from eq. (1.4), assuming all IT heat reaches the water
Rack airflow (rack at 20 kW)	$\dot{m}_{\text{air}}$	1.54 kg/s	calculated from eq. (1.3), assuming all rack heat leaves in the measured air stream

the whole 20 kW leaves the rack in the air stream whose inlet and outlet were measured, and that the two sensors represent the average temperatures of that stream. An *inferred* statement is a conclusion about something no sensor observes, reached by weighing the measurements against models of what could have produced them: "the airflow through this rack is probably inadequate", or

$$\Pr(\text{temperature sensor biased} \mid \text{data}) = 0.65.$$

An inferred statement carries its uncertainty with it, and is revised when new data arrive.

Keeping the three apart is a discipline the rest of the book enforces. The chilled-water flow of 47.8 kg/s in Table 1.2 looks like a fact, but it is a calculation that assumes the water carries all the IT heat; if a flow meter on the loop reads 40 kg/s, the disagreement does not show which of the two is wrong. It shows that one of the meters, or the assumption, is wrong, and choosing among those explanations is an inference.

1.6 Diagnosis as ranked hypotheses and the picture to keep

The coupling of networks, the gaps in measurement and the uncertainty in capacity make a data center a natural object for probabilistic reasoning. This section shows how a single alarm becomes a ranked list of explanations and a choice of the next measurement, then assembles the chapter's picture of the facility, contrasts it with a battery, and states the takeaway.

1.6.1 One symptom, six hypotheses

Suppose an alarm reports that a rack's inlet temperature is high: it reads 29°C where it normally reads about 24°C. Several physical causes produce that symptom, and Table 1.3 lists six. The first column names the hypothesis, the second describes its mechanism, and the third gives the signature it would leave in other measurements, which is what makes the hypotheses distinguishable.

Table 1.3: Six hypotheses for a high rack inlet temperature, with the mechanism of each and the evidence that would distinguish it.

Hypothesis	Mechanism	Signature in other measurements
H_1 higher IT load	rack and neighbors draw more power, so more exhaust and more recirculation	rack PDU power rises; outlet temperature rises
H_2 reduced airflow	less supply air reaches the rack (blocked tiles, a slowed or failed CRAH fan)	low supply pressure or tile airflow at the rack; nearby racks also warm
H_3 air recirculation	exhaust returns to the intake around the row end or over the rack (missing blanking panels, gaps)	supply airflow normal; inlet hottest at the top of the rack or the end of the row
H_4 CRAH degradation	coil fouled or valve faulty, so supply air leaves warmer	CRAH supply-air temperature rises; all racks served by that unit warm
H_5 chilled-water problem	warm or insufficient chilled water at the coils	chilled-water supply temperature rises or loop flow falls
H_6 bad sensor	the inlet sensor is biased or has failed	neighboring sensors and a portable reference read normal

An operator facing the alarm wants the probability of each hypothesis given the data y, $\Pr(H_i \mid y)$. By Bayes' rule it combines how plausible each hypothesis was before the alarm,

$\Pr(H_i)$, with how well each explains what was observed, $\Pr(y \mid H_i)$:

$$\Pr(H_i \mid y) = \frac{\Pr(y \mid H_i)\,\Pr(H_i)}{\sum_j \Pr(y \mid H_j)\,\Pr(H_j)}. \tag{1.10}$$

An illustrative result is the posterior plotted in Figure 1.7: air recirculation 42 %, reduced airflow 31 %, sensor bias 15 %, higher IT load 8 %, CRAH problem 4 %, and a chilled-water problem close to 0 %. The ranking follows from the other readings: the chilled-water supply is on its design value of 7°C, which nearly rules out H_5; the CRAH supply air is normal, which makes H_4 unlikely; the rack's PDU power has risen only slightly, which leaves H_1 a minor share; and the inlet sensors of the neighboring racks are only partly elevated, which keeps a sensor fault in contention. The horizontal axis of the figure lists the six hypotheses in order of probability, and the vertical axis is the posterior probability in percent.

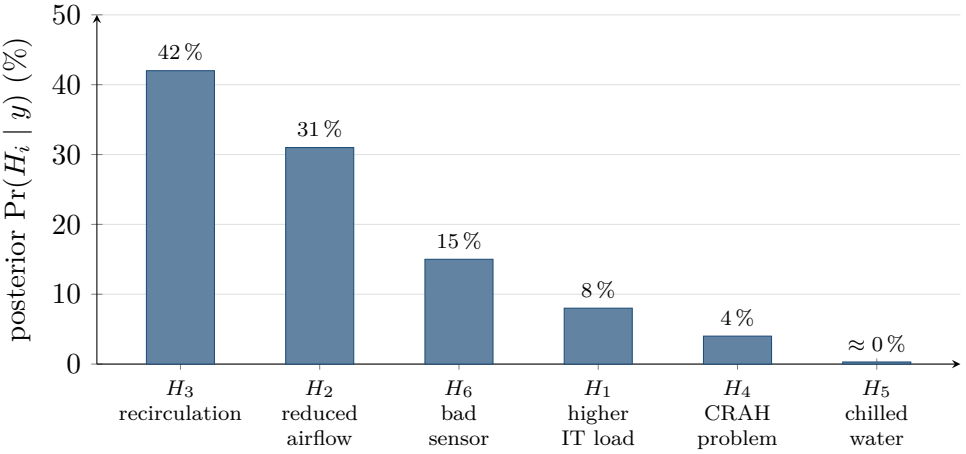


Figure 1.7: Illustrative posterior probabilities of six hypotheses for a high rack inlet temperature, ranked. The two leading hypotheses, recirculation and reduced airflow, together hold 73 % of the probability, and the next measurement is chosen to separate them.

1.6.2 The next measurement

The posterior answers "what is probably wrong"; the operator also needs "what to measure next". The two leading hypotheses differ in one clear signature (Table 1.3): reduced airflow predicts a low supply of air at the rack, recirculation predicts a normal supply with exhaust reaching the intake. A measurement of the supply airflow at the rack, for example the pressure difference between the raised-floor plenum and the cold aisle at the rack's perforated tiles, separates them. Suppose, illustratively, that this reading comes out low with probability 0.85 under H_2, 0.15 under H_3, 0.30 under H_4 (a CRAH fan fault lowers the pressure) and 0.10 under each of the others.

Table 1.4 applies equation (1.10) to both possible outcomes. Its columns are the current posterior, the probability of a low reading under each hypothesis, and the posterior after a low and after a normal reading. A low reading is expected with probability $\sum_j \Pr(\text{low} \mid H_j)\,\Pr(H_j) = 0.36$ and a normal one with probability 0.64. Either outcome reorders the leading pair decisively: a low reading moves reduced airflow from 31 to 73 %; a normal reading moves recirculation from 42 to 56 % and reduced airflow down to 7 %. The ratio of the two leading probabilities, $42/31 \approx 1.35$ before the measurement, becomes 0.24 or 7.7 after it. That is what makes the measurement worth taking: its value lies in how much it is expected to change the ranking. A measurement that both hypotheses predict equally, such as the chilled-water temperatures, would leave the posterior where it is, however precise it was. The sensor-bias

hypothesis, which this measurement barely touches, is then addressed by a different one, a portable reference thermometer at the rack inlet.

Table 1.4: Bayesian update of the illustrative posterior by a measurement of the supply airflow at the rack, for both possible outcomes.

Hypothesis	Current posterior	$\Pr(\text{low} \mid H_i)$	After a low reading	After a normal reading
H_3 recirculation	0.42	0.15	0.174	0.559
H_2 reduced airflow	0.31	0.85	0.729	0.073
H_6 bad sensor	0.15	0.10	0.041	0.211
H_1 higher IT load	0.08	0.10	0.022	0.113
H_4 CRAH problem	0.04	0.30	0.033	0.044
H_5 chilled water	≈ 0	0.10	≈ 0	≈ 0
Probability of the outcome			0.3615	0.6385

1.6.3 The system diagram

The chapter's picture of the facility can now be drawn in one diagram. Figure 1.8 places the energy chain in a single row: grid, UPS and power distribution, servers (where P_{IT} becomes $\dot{Q}_{\text{IT}}$), air, CRAH or CRAC, chilled water, chiller, and cooling tower or outdoors. Under each box are the illustrative values of the chapter at that stage. The band above the row is the sensing and control layer, with dashed arrows to the equipment it measures and adjusts. The notes below the row are the redundancy at each stage, and the dashed gold line along the bottom is the electrical supply of the cooling equipment, which closes the loop between the thermal and electrical networks.

grid $\rightarrow$ UPS and distribution $\rightarrow$ servers ($P_{\text{IT}} \rightarrow \dot{Q}_{\text{IT}}$) $\rightarrow$ air $\rightarrow$ CRAH/CRAC $\rightarrow$ chilled water $\rightarrow$ chiller $\rightarrow$ cooling tower or outdoors,

overlaid by sensors, controls and redundancy, is the diagram to keep.

1.6.4 Contrast with a battery

A comparison with a battery energy storage system shows what kind of physics dominates a data center. A battery stores a significant amount of energy: electrical power flows in and out, the stored energy rises and falls, and a part of the throughput becomes heat,

$$P_{\text{electrical}} \leftrightarrow E_{\text{stored}} \rightarrow \dot{Q}_{\text{thermal}}.$$

A data center stores almost nothing. Power flows in, the IT equipment consumes it, and the heat flows out,

$$P_{\text{electrical}} \rightarrow P_{\text{IT}} \rightarrow \dot{Q}_{\text{thermal}},$$

with no significant storage apart from short-term thermal storage in the masses of the equipment and building and in the air and water of the cooling loops. The key physics of a data center is therefore transport, capacity and transient storage: how heat moves, how much each path can carry, and how quickly temperatures change when the flows do not balance.

Thermal capacitance still matters, through the heat balance of each mass,

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}}, \tag{1.11}$$

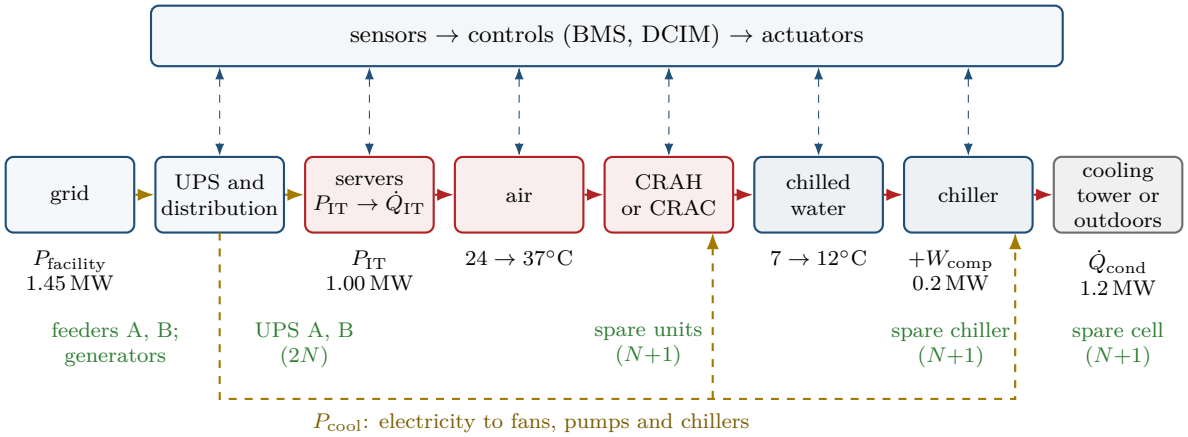


Figure 1.8: The system diagram of a data center: the energy chain from grid to outdoors (gold arrows electrical, red arrows heat), with the chapter's illustrative values under each stage, the redundancy at each stage (green), the sensing and control layer (top, dashed arrows both ways: measurements up, commands down) and the electrical supply of the cooling equipment (dashed gold line).

but it sets how quickly temperatures change, not how much useful operational energy is stored. Two illustrative estimates show the scale. The air of a $25 \times 20 \times 4\,\text{m}$ data hall, $2000\,\text{m}^3$ at $1.2\,\text{kg/m}^3$, has a thermal capacitance of $2400\,\text{kg} \times 1.0\,\text{kJ/(kg K)} = 2.4\,\text{MJ/K}$; if all cooling stopped under a 1 MW IT load, the air alone would warm at $10^6/(2.4 \times 10^6) \approx 0.42\,\text{K}$ per second, about 25 K per minute. The equipment, floor and walls add capacitance and slow this, but only to minutes. The water of a chilled-water loop of $100\,\text{m}^3$ has $10^5\,\text{kg} \times 4.18\,\text{kJ/(kg K)} = 418\,\text{MJ/K}$; under the same 1 MW imbalance it warms at about 0.0024 K per second, 0.14 K per minute, so a 5 K margin lasts about 35 minutes. That margin holds $418\,\text{MJ/K} \times 5\,\text{K} \approx 2.1\,\text{GJ} \approx 0.58\,\text{MWh}$ of cooling, enough to ride through a chiller restart, which is why some plants add chilled-water storage tanks, but it is not energy the facility can use to run its servers. A battery stores electrical energy for later use; a data center's thermal masses only delay the consequences of an imbalance.

1.6.5 A coupled engineered system

The chapter's conclusion is that a data center is not a room of computers. It is a coupled engineered system of electrical, IT, thermal, fluid, control and reliability subsystems, linked by conservation laws: conservation of energy along the power path and through the IT equipment, conservation of mass and energy in the air and water loops, and the energy balance of the chiller that adds its work to the heat it moves. The central chain is

$$\text{electricity} \rightarrow \text{computation} \rightarrow \text{heat} \rightarrow \text{air and water} \rightarrow \text{environment},$$

and every later chapter refines one link of it or the layers that measure, control and protect it.

The central operational question follows from the chain: whether the IT equipment can be kept powered and within safe physical limits despite changing loads and equipment failures. Each part of the question corresponds to a part of the system. "Powered" is the electrical network and its redundant paths; "within safe physical limits" is the thermal-fluid network and its capacity; "despite changing loads and failures" is the control loop, the redundancy levels and the probability of inadequacy; and answering it at all requires separating what is measured from what is calculated and what must be inferred.

1.7 What to retain

Definition 1.1 (Data center as a coupled system)**.** A data center is a coupled engineered system, electrical, IT, thermal, fluid, control and reliability, linked by conservation laws along the chain electricity → computation → heat → air and water → environment. Its central operational question is whether the IT equipment can be kept powered and within safe physical limits despite changing loads and failures.

Table 1.5 lists the chapter's objects and the subsection that defines each.

Table 1.5: The objects of the system architecture and where each is defined.

Object	Expression	Role	Defined in
Three layers	Figure 1.1	energy chain, control, resilience	1.1.1
IT heat	$\dot{Q}_{\mathrm{IT}} \approx P_{\mathrm{IT}}$	all IT power becomes heat	1.1.2
Power path	grid to server	delivery and continuity	1.2.1
Rack heat	$\dot{Q}_{\mathrm{rack}} \approx P_{\mathrm{rack}}$	heat concentrated per cabinet	1.2.3
Air heat balance	$\dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} (T_{\mathrm{out}} - T_{\mathrm{in}})$	first transport medium	1.3.1
Hot and cold aisles	Figure 1.3	separate supply from exhaust	1.3.2
Water heat balance	$\dot{m}_w c_{p,w} (T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}})$	transport to the plant	1.3.4
Heat-flow chain	equation (1.5)	air side equals water side	1.3.4
Chiller balance	$\dot{Q}_{\mathrm{cond}} = \dot{Q}_{\mathrm{evap}} + W_{\mathrm{comp}}$	lifts heat, adds work	1.3.5
Loop hierarchy	Figure 1.4	chip to atmosphere	1.3.6
Three networks	electrical, thermal-fluid, control	coupled through rack and P_{cool}	1.4.1
State form	$\dot{x} = f(x, u, d, \theta)$	states, inputs, disturbances, parameters	1.4.2
Facility power	equation (1.8)	IT plus overhead	1.4.3
N, $N{+}1$, $2N$	Table 1.1	redundancy levels	1.5.1
Adequacy risk	$\Pr(\dot{Q}_{\mathrm{avail}} < \dot{Q}_{\mathrm{req}})$	capacity as a probability	1.5.2
Measured, calculated, inferred	Table 1.2	three kinds of statement	1.5.4
Ranked hypotheses	$\Pr(H_i \mid y)$	diagnosis of one symptom	1.6.1
Thermal capacitance	$C_{\mathrm{th}}\, \mathrm{d}T/\mathrm{d}t = \dot{Q}_{\mathrm{in}} - \dot{Q}_{\mathrm{out}}$	sets speed, stores little	1.6.4

1.8 Exercises

Exercise 1.1. A rack draws $P_{\mathrm{rack}} = 30\,\mathrm{kW}$. Its inlet air is at 23°C and its outlet at 35°C. (a) Compute the airflow through the rack in kg/s and in m^3/s (air density $1.2\,\mathrm{kg/m^3}$). (b) A hall of such racks has $P_{\mathrm{IT}} = 1.5\,\mathrm{MW}$ and a chilled-water loop with supply at 8°C and return at 14°C. Compute the chilled-water flow, assuming all the IT heat reaches the water. (c) The chillers run at COP = 5. Compute the compressor power and the heat the cooling towers must reject.

Exercise 1.2. A facility requires 4 MW of cooling and uses chillers of 1.5 MW each. (a) How many chillers does each of the designs N, $N{+}1$ and $2N$ install, and what capacity remains after one failure in each? (b) Each chiller is available independently with probability 0.97. Compute the probability that the $N{+}1$ plant falls short of 4 MW. (c) On hot days each chiller delivers

only 1.3 MW. Recompute (b) for a hot day, and explain why the $N{+}1$ label does not change while the risk does.

Exercise 1.3. A facility reports $P_{\text{facility}} = 2.2\,\text{MW}$, $P_{\text{IT}} = 1.6\,\text{MW}$, a rack power of 26 kW, a rack inlet temperature of 25°C and an outlet temperature of 38°C. (a) Compute the facility overhead, the ratio $P_{\text{facility}}/P_{\text{IT}}$ and the rack airflow implied by the readings. (b) Classify each of the following as measured, calculated or inferred, and state the assumption behind each calculated one: the inlet temperature; the rack airflow from (a); "the rack's airflow is probably below the 2.6 kg/s its servers are designed to draw". (c) Name two quantities in this facility that none of the readings determines.

Exercise 1.4. Start from the posterior of Figure 1.7 (H_3 0.42, H_2 0.31, H_6 0.15, H_1 0.08, H_4 0.04, H_5 0). A portable reference thermometer is placed at the rack inlet. It agrees with the installed sensor with probability 0.05 if the sensor is biased (H_6) and with probability 0.95 under every other hypothesis. (a) Compute the probability that the two agree. (b) Compute the posterior of all hypotheses if they agree, and if they disagree. (c) Compare the effect of this measurement with that of the airflow measurement of Table 1.4: which hypothesis does each one mainly resolve?

1.9 Connection to the TERRA framework

The architecture of this chapter is the graph a TERRA model of a data center is built on.

- **Three coupled conservation networks.** The electrical network of Figure 1.5 is a power-conservation graph whose nodes are buses (switchgear, UPS output, PDU) and whose edges carry power; the thermal-fluid network is a mass graph for air and water and an energy graph for heat, with nodes at plenums, aisles, racks, coils and loops. The rack node couples them through $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$, and the cooling equipment couples them back through P_{cool} in equation (1.8).
- **Components with efficiencies, capacities and bounds.** UPS, coil, chiller, fan, pump and tower are edge components: the chiller relates $\dot{Q}_{\text{evap}}$, $\dot{Q}_{\text{cond}}$ and W_{comp} through equation (1.6) and a COP, each has a capacity, and each redundant unit of Table 1.1 is a parallel edge that may be absent.
- **Balances that should close are residuals.** The chain of equation (1.5), IT power against air-side heat against water-side heat, and the chiller balance, evaporator plus work against condenser, are redundancy relations: computed from independent meters, their mismatches are the residuals from which biased sensors and unmetered heat paths are detected.
- **Measured, calculated, inferred.** The readings of Table 1.2 are measurements y in equation (1.7); rack airflow, recirculation fraction α and coil conductance UA are the unmeasured states and parameters x and θ that the framework infers as posterior distributions, and which-rack-is-on-which-path is itself an uncertain part of the graph.
- **Risk and diagnosis as probabilities.** Adequacy is computed as $\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}})$ by forward simulation over uncertain load, weather and unit availability; an alarm becomes a ranked posterior over hypotheses H_1 to H_6 (Figure 1.7), and the next measurement is the one whose outcomes are expected to change that posterior most (Table 1.4).

In summary, a data center is three coupled networks, electrical, thermal-fluid and information, through which energy flows from the grid to the atmosphere with almost no storage, protected

by parallel paths and watched by sensors that observe only part of the state; its operation is the question of whether capacity stays above requirement, and its diagnosis is a ranking of hypotheses by probability. The next chapter starts at the head of the thermal chain, the IT load itself: how much power servers draw and how that power varies with their workload.

Chapter 2

IT loads and computing demand: from workload to electrical power and heat

2.1 Server power

A data center exists to run its information technology (IT) equipment: the servers that compute, the storage systems that hold data and the network equipment that moves it. Every other system in the facility, the electrical distribution, the backup power, the air handlers, the chillers and the cooling towers, is there to supply that equipment with power and to remove the heat it produces. The IT equipment is therefore the load that sizes everything else, and the question of this chapter is

what determines the electrical power that IT equipment consumes, and how a demand for computation turns into electrical load and into heat.

The answer is built in layers. This section starts with one server and the components whose power adds up to its consumption; the following sections aggregate servers into racks, connect the number and activity of the servers to the computing demand they serve, show how the load can be moved and shaped, follow it in time and through the rest of the facility, and close with what can be measured of it.

Jargon: utilization, idle power, dynamic and static power, rack density, power cap, scheduler

Utilization u is the fraction of a processor's or a server's capacity that is in use over a sampling window, between 0 (no work) and 1 (fully busy). For a central processing unit (CPU) it is usually the fraction of time the cores are executing instructions rather than waiting; for a graphics processing unit (GPU) or another accelerator it is usually the fraction of time at least one computation is running on it. *Idle power* P_{idle} is the power a server draws when it is switched on and ready but doing no useful work; the power it draws at full load is $P^{\max}$. *Dynamic power* is the part of a chip's power that is spent switching transistors and therefore scales with activity; *static power* (leakage power) is the part that flows through the transistors even when they do not switch. *Rack density* is the electrical power consumed by the equipment in one rack, the standard cabinet about 0.6 m wide and 1.0 to 1.2 m deep that holds servers, usually stated in kilowatts per rack. A *power cap* is an upper limit on the power a server, a processor or a group of servers may draw, enforced by the server's firmware, which lowers clock frequency and voltage to stay under it. A *scheduler* is the software that decides which jobs or requests run on which servers and when, and therefore how many servers are active and how busy each one is.

Key idea

IT load is not fixed. It is set by the computing demand, by the scheduler's decisions about how many servers run and how busy they are, and by the physics of the processors; and because nearly all of it becomes heat, every change in computing demand is also a change in the cooling demand of the facility.

2.1.1 Components of server power

A server is an assembly of components that each draw power from its power supply: one or more CPUs, possibly GPUs or other accelerators, memory, storage drives, network interfaces, cooling fans, and the power supply itself, which loses part of the power it converts. The power the server draws from its electrical feed is the sum

$$P_{\text{server}} = P_{\text{CPU}} + P_{\text{GPU}} + P_{\text{mem}} + P_{\text{stor}} + P_{\text{net}} + P_{\text{fan}} + P_{\text{PSU}}, \tag{2.1}$$

where the subscripts denote the processors, the accelerators, the memory, the storage, the network interfaces, the fans and the losses of the power supply. Equation (2.1) is a statement of energy conservation at the server's electrical input: every watt drawn from the feed is consumed by one of the components or dissipated in the conversion from the feed's alternating current to the low-voltage direct current the components use.

The terms do not vary equally. Network interfaces and storage drives draw nearly the same power whether the server is busy or not; memory varies moderately with how often it is accessed; fans speed up as the components heat; the power-supply loss grows with the power converted. The dominant variation comes from the processors: the CPUs, the GPUs and other accelerators, and sometimes the memory. In a server built around accelerators the GPUs dominate both the level and the variation, and a single modern accelerator can draw several hundred watts by itself.

Figure 2.1 shows the decomposition for an illustrative CPU server without accelerators. The horizontal axis has two operating points, idle ($u = 0$) and full load ($u = 1$); the vertical axis is the server power in watts, and each bar is stacked by component in the order of the legend, the CPU at the bottom. The server draws 200 W idle and 600 W at full load. Of the 400 W difference, 250 W is the CPU; memory, fans and power-supply losses account for most of the rest, and storage and network barely change. The power-supply losses at full load, 65 W

of 600 W, correspond to a conversion efficiency near 89 % in this illustration. Fan power also depends on the temperature of the air entering the server, because the server's firmware raises fan speed to hold its component temperatures: the same computation in a warmer room draws more fan power, so IT power is not determined by computation alone.

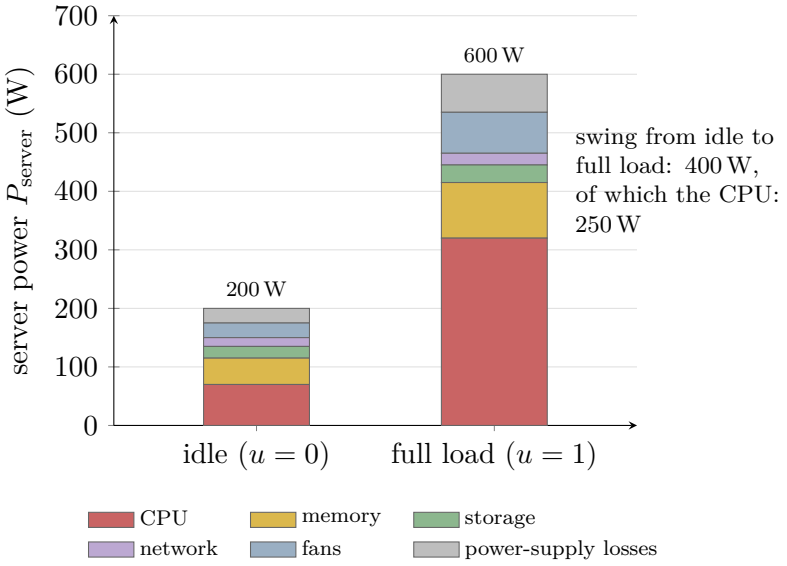


Figure 2.1: Power of an illustrative CPU server by component, equation (2.1), at idle and at full load. Values in watts: CPU 70 and 320, memory 45 and 95, storage 20 and 30, network 15 and 20, fans 25 and 70, power-supply losses 25 and 65. The processor carries most of the variation; the idle bar is far from zero.

2.1.2 Idle power and the linear power model

The idle bar of Figure 2.1 is the first property of IT load: a server that is switched on and doing nothing still consumes a substantial fraction of its full-load power. The illustrative server draws $P_{\text{idle}} = 200\,\text{W}$ idle and $P^{\max} = 600\,\text{W}$ at full load, so a third of its maximum is spent before any work is done. Memory refresh, disk spin or controller power, network links that must stay up, fans at minimum speed, the power supply's own losses and the leakage of the processors all persist at zero utilization.

The simplest model that respects this is linear in utilization between the two end points,

$$P_{\text{server}}(u) = P_{\text{idle}} + u\,(P^{\max} - P_{\text{idle}}), \qquad 0 \le u \le 1. \tag{2.2}$$

The intercept is the idle power and the slope is the dynamic range $P^{\max} - P_{\text{idle}}$. For the illustrative server at half utilization,

$$P_{\text{server}}(0.5) = 200 + 0.5 \times (600 - 200) = 200 + 0.5 \times 400 = 400\,\text{W}.$$

Half the work costs two thirds of the full-load power, not half of it, because the idle part is paid in full whatever the utilization. A server whose power were proportional to utilization, $P_{\text{server}} = u\,P^{\max}$, would draw 300 W at $u = 0.5$; the 100 W difference is the idle power scaled by the unused fraction, $(1 - u)\,P_{\text{idle}} = 0.5 \times 200$.

Figure 2.2 plots equation (2.2). The horizontal axis is utilization from 0 to 1; the vertical axis is the server power in watts. The solid line is the linear model; its intercept at $u = 0$ is the idle power, marked by the brace on the left, and the brace on the right marks the dynamic range. The dashed curve is an illustrative non-linear relation of the kind measured

on real servers, $P_{\text{server}} = P_{\text{idle}} + (P^{\max} - P_{\text{idle}})\, u^{0.7}$, which rises faster at low utilization and more slowly near full load. At $u = 0.5$ the linear model gives 400 W and the non-linear curve $200 + 400 \times 0.5^{0.7} = 446$ W, a difference of about 10 %. The linear model is crude, but it is useful: it has two parameters, both of which can be read from a server's specification or estimated from a few paired measurements of power and utilization, and it captures the property that matters most for operation, the idle offset.

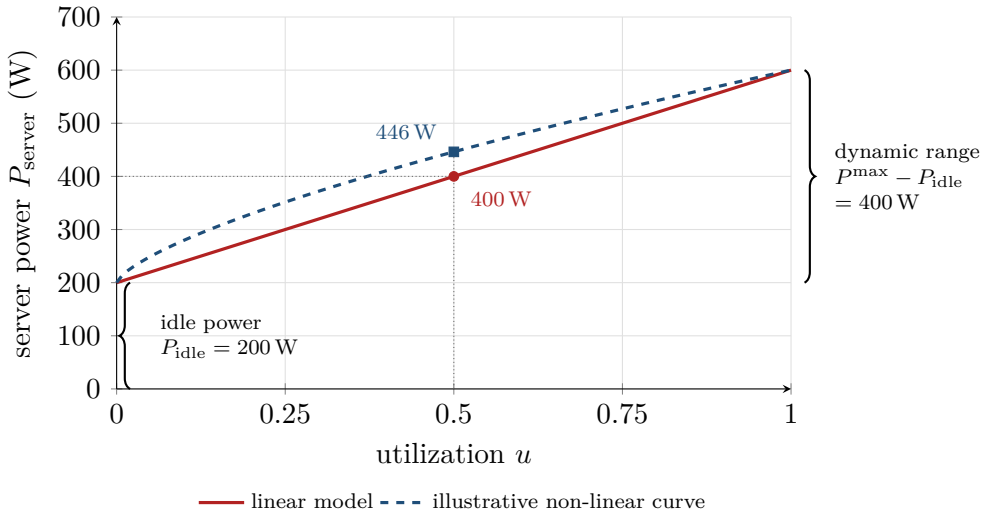


Figure 2.2: Server power against utilization for the illustrative server with $P_{\text{idle}} = 200$ W and $P^{\max} = 600$ W. Solid: the linear model $P_{\text{server}} = P_{\text{idle}} + u(P^{\max} - P_{\text{idle}})$. Dashed: an illustrative non-linear relation $P_{\text{server}} = P_{\text{idle}} + (P^{\max} - P_{\text{idle}})\, u^{0.7}$. At $u = 0.5$ the two give 400 and 446 W.

Principle 2.1. *A switched-on server draws its idle power whatever work it does. Power is affine in utilization, not proportional to it, so half the work costs more than half the full-load power.*

2.1.3 Utilization is not power

The linear model treats utilization as the single driver of power. It is not. A CPU at 50 % utilization does not necessarily draw half its maximum power, nor the power that equation (2.2) predicts, because utilization measures how much of the time the processor is busy, not how hard it works while busy. The power drawn while busy depends on the clock frequency at which the processor runs, the supply voltage, the type of instruction being executed (simple integer operations or wide vector and matrix operations), how often the processor reaches out to memory and how well its caches serve those accesses, whether the chip is throttling its clock to stay within a temperature limit, and how many of its cores are active. In general

$$P_{\text{CPU}} = f(u,\ f_{\text{clk}},\ V,\ \text{workload type},\ T,\ \ldots), \tag{2.3}$$

where f_{clk} is the clock frequency, V the supply voltage and T the chip temperature. Utilization is one argument among several.

The gap is wider for GPUs. The utilization that GPU monitoring tools report is commonly the fraction of time at least one computation kernel is running on the device, not the fraction of its arithmetic units that are busy. Two jobs that each show 90 % GPU utilization can therefore draw noticeably different power: a dense matrix-multiplication job in the training of a large neural network keeps nearly every arithmetic unit busy every cycle, while a job limited by

memory bandwidth keeps the device occupied but leaves most of its arithmetic idle. With illustrative numbers, the first might draw 400 W and the second 250 W at the same reported 90 %. A utilization counter is a measurement of occupancy; power is a measurement of energy flow, and the two are related only through a model that knows the workload.

2.1.4 Dynamic and static power

The physics of a processor explains both the idle offset and the dependence on frequency and voltage. Processors are built in complementary metal-oxide-semiconductor (CMOS) logic. Each time a logic gate switches, it charges or discharges a small capacitance through the transistors, and the energy taken from the supply per charge and discharge cycle is proportional to that capacitance times the square of the voltage. Summed over all gates and all clock cycles per second, the switching or dynamic power is

$$P_{\text{dyn}} \propto C_{\text{eff}}\, V^2 f_{\text{clk}}, \tag{2.4}$$

where C_{eff} is the effective switching capacitance, the capacitance switched per cycle, which includes how many of the gates actually change state and therefore depends on the instructions being executed; V is the supply voltage; and f_{clk} is the clock frequency. The dependence on instructions is the physical reason for the two 90 % GPU jobs of the previous subsection: they switch different numbers of gates per cycle.

Voltage and frequency are not independent. A transistor switches faster at a higher voltage, so running the clock faster usually requires raising the voltage as well; processors adjust both together, a technique called dynamic voltage and frequency scaling. Over the usual operating range the voltage rises roughly in proportion to frequency, and equation (2.4) then makes dynamic power grow roughly as the cube of frequency. With illustrative numbers, raising the clock by 20 % with a 10 % higher voltage multiplies dynamic power by $1.20 \times 1.10^2 = 1.45$: 20 % more speed for 45 % more power. Read in the other direction, a modest reduction in frequency saves a disproportionate amount of power, which is what makes power capping (Section 2.4.3) cheap in performance.

The static or leakage power is the current that flows through transistors that are nominally off, multiplied by the supply voltage. It does not depend on switching activity, so it persists when the processor is idle, and it rises with chip temperature. The total is

$$P_{\text{CPU}} = P_{\text{dyn}} + P_{\text{static}}, \tag{2.5}$$

and $P_{\text{CPU}} \neq 0$ even at very low activity. The static term is one of the contributors to the idle power of Section 2.1.2, and its dependence on temperature is a second coupling between the thermal environment and the electrical load, after the fans: a hotter chip leaks more and draws more power for the same work.

2.2 Rack power and rack density

Servers are installed in racks, and the rack is the unit at which a data center plans its electrical circuits and its cooling. This section aggregates server power to rack power, shows that rack power and rack heat are the same number, and explains why the power density of a rack determines how the rack must be cooled.

2.2.1 From server to rack

A rack holds tens of servers together with the equipment they share: a top-of-rack network switch, sometimes storage enclosures, and the rack power distribution units (PDUs) that split

the rack's feed among the servers and usually meter it. The rack's power is the sum over its servers plus that shared equipment and the distribution losses inside the rack,

$$P_{\text{rack}} = \sum_{i=1}^{n} P_{\text{server},i} + P_{\text{shared}} + P_{\text{dist}}, \tag{2.6}$$

For an illustrative rack of 40 servers each drawing 500 W, the servers alone draw

$$40 \times 500\,\text{W} = 20\,000\,\text{W} = 20\,\text{kW},$$

and with networking, storage and losses adding about 2 kW the rack draws $P_{\text{rack}} \approx 22\,\text{kW}$.

The same rack produces about 22 kW of heat. The electrical energy that enters a server does no lasting work on its surroundings: it is not lifted, not stored chemically, not sent out of the room except as a negligible energy of network signals and light. Every joule that enters is converted to heat in the chips, the memory, the power supply and the fan motors, and leaves in the air or the liquid that cools the equipment. In steady operation, energy conservation for the rack as a control volume gives

$$\dot{Q}_{\text{rack}} \approx P_{\text{rack}}, \tag{2.7}$$

and for the whole IT load $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$. The relation is the first link between the electrical and the thermal sides of a data center: the rack power, which is measured by the rack's PDUs, gives the rack heat by calculation, without any thermal measurement. The heat balance on which this rests is the subject of Chapter 5.

Principle 2.2. *In steady operation the heat a rack releases equals the electrical power it consumes. Every kilowatt of IT power is a kilowatt of cooling demand.*

2.2.2 Rack density

The industry describes racks by their power: a "10 kW rack", a "30 kW rack", an "80 kW rack", a "120 kW rack". The number is the electrical power consumed in the rack, and by equation (2.7) it is simultaneously the heat that must be removed from the rack's footprint, so a rack density is at the same time an electrical rating and a cooling problem.

Table 2.1 makes the cooling problem quantitative for the four densities. Its columns are the rack density, the air mass flow needed to carry that heat away with an illustrative air temperature rise of 12 K through the servers, the corresponding volumetric flow, the heat released per square meter of rack footprint for an illustrative footprint of $0.6 \times 1.2 = 0.72\,\text{m}^2$, and the cooling approach indicatively associated with each class. The air flow follows from the heat balance of air developed in Chapter 5, $\dot{Q} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, \Delta T$, with $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$:

$$\dot{m}_{\text{air}} = \frac{\dot{Q}_{\text{rack}}}{c_{p,\text{air}}\,\Delta T} = \frac{80\,\text{kW}}{1.0\,\text{kJ/(kg K)} \times 12\,\text{K}} = 6.7\,\text{kg/s}$$

for the 80 kW rack, about $5.6\,\text{m}^3/\text{s}$ at an air density of $1.2\,\text{kg/m}^3$. The air flow grows in proportion to the rack power, and at 80 or 120 kW it is no longer practical to push that much air through one cabinet; the heat flux per footprint, above $100\,\text{kW/m}^2$, is several times that of a conventional room.

An 80 kW rack places in less than a square meter the heat of tens of household space heaters at full output, continuously. This is why high-density computing for artificial intelligence changes the cooling architecture of a data center rather than only its size: beyond a few tens of kilowatts per rack the heat must be collected by a liquid close to the chips, because water carries about 3500 times more heat per unit volume than air for the same temperature rise. The air side of the problem is the subject of Chapter 6 and the liquid side of Chapter 7.

Table 2.1: Rack density classes and the air flow needed to cool them, for an illustrative 12 K air temperature rise and a 0.72 m^2 footprint. The cooling approach is indicative; the boundaries between classes vary between designs.

Rack density	Air flow (kg/s)	Air flow (m^3/s)	Heat per footprint (kW/m^2)	Indicative cooling approach
10 kW	0.83	0.69	14	room air cooling
30 kW	2.5	2.1	42	air with aisle containment or rear-door heat exchangers
80 kW	6.7	5.6	111	liquid cooling at the chip, air for the remainder
120 kW	10	8.3	167	liquid cooling at the chip or immersion

2.2.3 Traditional and accelerator-oriented loads

The load a rack carries depends on the computing it does. Traditional enterprise workloads, databases, web servers, storage, business applications and virtualized servers (many virtual machines sharing one physical server), run on CPU servers with moderate power each, and give moderate rack densities, conceptually in the range of 10 to 20 kW per rack. Their load varies with the business day and with the number of users, but no single job moves it much.

Accelerator-heavy racks, built for training and running large machine learning models, hold servers with four or eight GPUs each, every GPU drawing several hundred watts, and reach many tens of kilowatts per rack or more. Their load is also more correlated: a training job spreads across hundreds or thousands of GPUs that compute in lockstep, so they start, pause and finish together, and the load of a whole hall can move by megawatts at once. The chain that the comparison exposes is

$$\text{compute architecture} \rightarrow \text{rack power density} \rightarrow \text{cooling architecture:}$$

the choice of hardware fixes the power per rack, and the power per rack fixes how the heat can be removed. A facility designed for the first kind of load cannot be filled with the second without rebuilding its cooling.

2.3 From computing demand to electrical load

The power of a server depends on its utilization, and its utilization depends on the work it is given. This section connects the demand for computation to the electrical load through the scheduler, names the levels of the connection, and gives the simplest quantitative model of it.

2.3.1 The demand-to-heat chain

When the demand for computation rises, more requests arrive or more jobs are submitted, the scheduler responds by placing work on more servers or GPUs, so the number of active servers N_{active} rises; the active servers become busier, so their utilization u rises; by equation (2.2) their power rises, so the IT power P_{IT} rises; and by equation (2.7) the heat $\dot{Q}_{\text{IT}}$ rises with it. Written as a chain,

$$D_{\text{compute}} \;\rightarrow\; N_{\text{active}},\; u \;\rightarrow\; P_{\text{IT}} \;\rightarrow\; \dot{Q}_{\text{IT}}, \tag{2.8}$$

where D_{compute} is the computing demand. Each arrow is a relation of a different kind. The first is a decision, made by the scheduler according to its policy; the second is the physics of the processors, Section 2.1; the third is energy conservation, Section 2.2.1.

Figure 2.3 draws the chain. The upper row of boxes is equation (2.8), from the computing demand on the left to the heat on the right, with the scheduler as an explicit stage. The middle row instantiates each stage for an accelerator cluster: the workload W, the number of GPUs assigned N_{GPU}, their utilization u_{GPU}, their power P_{GPU}, and the heat released in the GPUs. The bottom row names the lever that acts at each stage, which is the subject of Section 2.4. The heat does not end the chain: it becomes the cooling demand that the plant of Chapter 7 must meet.

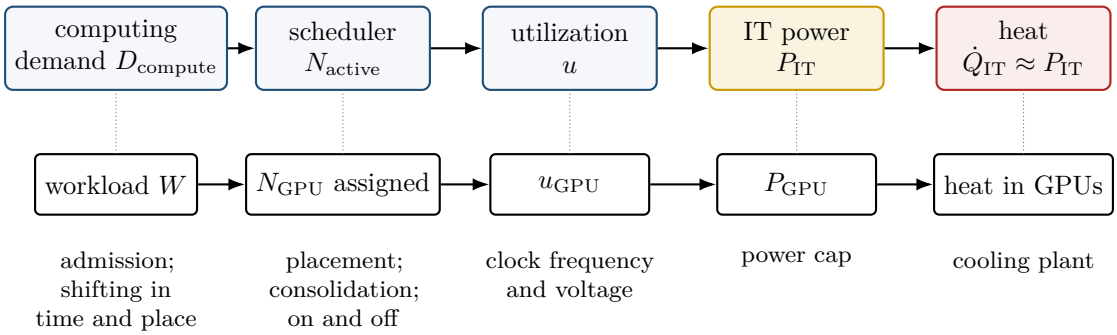


Figure 2.3: The chain from computing demand to heat, equation (2.8). Upper row: the general stages. Middle row: the same stages for an accelerator cluster. Bottom row: the lever that acts at each stage. A decision (the scheduler), processor physics and energy conservation connect the stages.

2.3.2 Levels of the chain

The chain has several levels, and they are worth keeping apart: the workload (what is being asked of the computers), the scheduler (which decides placement and timing), the active hardware (which servers and accelerators are switched on and assigned), the utilization of that hardware, and its electrical power. For an accelerator cluster,

$$W \;\rightarrow\; N_{\mathrm{GPU}} \;\rightarrow\; u_{\mathrm{GPU}} \;\rightarrow\; P_{\mathrm{GPU}}.$$

An observer who sees only the electrical power sees the last level; the earlier levels are visible only in the computing systems' own records.

The levels matter as soon as the load is to be controlled rather than only measured, because each level offers a different lever with a different cost and speed, as the bottom row of Figure 2.3 lists. Refusing or delaying work acts at the workload level and changes what gets computed; consolidating work acts at the hardware level and changes how many servers carry idle power; lowering clock frequency acts at the processor level and changes how fast the work completes; a power cap acts directly on the power and lets the processor choose its frequency to respect it. The same reduction in P_{IT} can be obtained at any of the levels, and the choice among them is a choice of which consequence (lost work, delay, slower service) is least costly.

2.3.3 An arrival-rate model

For a service that answers requests, such as a web application, the link from demand to hardware can be made quantitative. Let requests arrive at a rate λ, in requests per second, and let one server handle μ requests per second at its planned operating point. The number of servers required is at least

$$N_{\mathrm{required}} \approx \frac{\lambda}{\mu}, \tag{2.9}$$

rounded up to a whole server. With illustrative numbers, $\lambda = 5000$ requests/s and $\mu = 500$ requests/s per server give

$$N_{\text{required}} \approx \frac{5000}{500} = 10 \text{ servers},$$

and if each active server draws 450 W at its operating point,

$$P_{\text{IT}} \approx 10 \times 450\,\text{W} = 4500\,\text{W} = 4.5\,\text{kW}.$$

The model reads demand → number of servers → power, which is equation (2.8) with numbers attached: a doubling of the arrival rate doubles the number of servers and, at the same operating point, the power.

Equation (2.9) is a lower bound. A server asked to handle requests at exactly its capacity has no margin for the random bunching of arrivals, and its queue, and with it the response time, grows without limit as the load approaches capacity; in the simplest queueing model the mean waiting time is proportional to $1/(1-\rho)$, where $\rho = \lambda/(N\mu)$ is the utilization of the servers. Operators therefore provision for a target utilization below one. With an illustrative target $\rho = 0.7$ the same demand needs $5000/(0.7 \times 500) = 14.3$, that is 15 servers, each at a utilization of $5000/(15 \times 500) = 0.67$. For the 200 to 600 W server of Section 2.1.2, the ten servers at full utilization would draw $10 \times 600 = 6.0$ kW, while the fifteen at two-thirds draw $15 \times (200 + 0.67 \times 400) = 15 \times 467 = 7.0$ kW: the margin for response time costs 1 kW, five servers' worth of idle power. Response time, power and the number of servers are traded against one another through the idle power.

2.4 Flexibility of the load

Because the IT load is the outcome of scheduling decisions, it can be shaped. This section describes the responses available when demand falls, works through the consolidation example that shows why placement matters, and then treats three forms of control: power capping, shifting work in time and shifting it between sites.

2.4.1 Responses to falling demand

When the computing demand drops, at night for an interactive service or at the end of a batch of jobs, an operator has several options:

- leave all servers on, each at low utilization;
- consolidate the work onto fewer servers, each at higher utilization;
- turn the servers left idle off, or put them in a low-power sleep state;
- shift work that can wait to a later time;
- shift work to another data center.

The first is what happens if nothing is decided. Consider two strategies for the same work. Strategy A runs 20 servers at $u = 0.25$; strategy B runs 5 servers at $u = 1$ and switches the other 15 off. If throughput is proportional to utilization, both deliver $20 \times 0.25 = 5 \times 1 = 5$ server-equivalents of work, the same nominal throughput. Their power is not the same, because every active server carries its idle power.

The linear model makes this exact. For N identical active servers with utilizations $u_1, \ldots, u_N$, summing equation (2.2) gives

$$P_{\text{group}} = N\,P_{\text{idle}} + (P^{\max} - P_{\text{idle}}) \sum_{i=1}^{N} u_i. \tag{2.10}$$

The sum of utilizations is the work done; for a fixed amount of work the second term is fixed, and the only term that depends on the placement is NP_{idle}. Workload placement therefore affects electrical efficiency: the fewer servers carry a given amount of work, the less idle power is paid for it.

2.4.2 A consolidation example

Take the illustrative server, $P_{\text{idle}} = 200\,\text{W}$ and $P^{\max} = 600\,\text{W}$. Under strategy A each of the twenty servers draws

$$200 + 0.25 \times 400 = 300\,\text{W}, \qquad 20 \times 300\,\text{W} = 6000\,\text{W} = 6\,\text{kW}.$$

Under strategy B each of the five servers draws

$$200 + 1.00 \times 400 = 600\,\text{W}, \qquad 5 \times 600\,\text{W} = 3000\,\text{W} = 3\,\text{kW}.$$

The same nominal computation costs 6 kW one way and 3 kW the other. By equation (2.10), both strategies spend $400\,\text{W} \times 5 = 2\,\text{kW}$ on dynamic power; strategy A spends $20 \times 200\,\text{W} = 4\,\text{kW}$ on idle power and strategy B $5 \times 200\,\text{W} = 1\,\text{kW}$.

Figure 2.4 draws the two strategies as two groups of bars. Each bar is one active server; its height is the server's power in watts, split into the idle part (gray, the same 200 W in every bar) and the dynamic part $u(P^{\max} - P_{\text{idle}})$ (red). The left group is strategy A, twenty short bars; the right group is strategy B, five tall bars. The red area is the same in both groups; the gray area of the left group is four times that of the right.

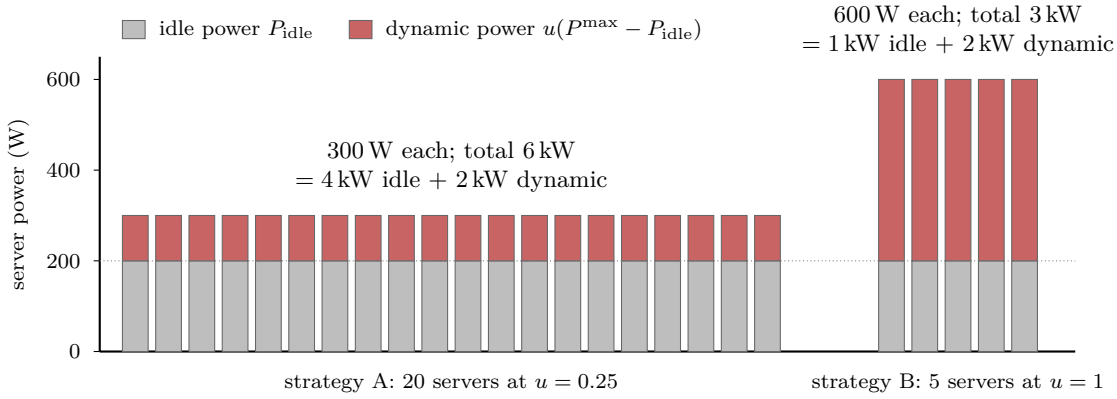


Figure 2.4: Consolidation with the linear model, equation (2.10), for the illustrative server ($P_{\text{idle}} = 200\,\text{W}$, $P^{\max} = 600\,\text{W}$). Both strategies deliver five server-equivalents of work and the same 2 kW of dynamic power; strategy A pays idle power on twenty servers, strategy B on five.

Two qualifications apply to the example. First, the saving exists only if the fifteen freed servers are switched off or put to sleep: left on and idle they draw $15 \times 200\,\text{W} = 3\,\text{kW}$, and strategy B costs $3 + 3 = 6\,\text{kW}$, exactly the same as strategy A. The saving of consolidation is the idle power of the servers it allows to be turned off. Second, running servers at $u = 1$ leaves no margin, and the queueing argument of Section 2.3.3 applies: an interactive service consolidated to full utilization answers slowly. A server that has been switched off also takes time to start again, from seconds for a sleep state to minutes for a full start. In practice operators consolidate toward a target utilization below one and keep some servers in reserve, and the attainable saving lies between the two strategies. The direction of the result does not change.

Principle 2.3. *For a fixed amount of work the dynamic power is fixed and the idle power is proportional to the number of active servers. Where the work is placed, and how many servers are left on, changes the electrical load.*

2.4.3 Power capping

A power cap imposes an upper limit on the power a server, a processor or a group of servers may draw,

$$P_{\text{server}} \leq P_{\text{cap}}, \tag{2.11}$$

for example $P_{\text{GPU}} \leq 350\,\text{W}$ for each GPU of a cluster. The firmware enforces the limit by lowering clock frequency and voltage until the power falls under it. The consequence is some loss of performance, the work takes longer, against a lower electrical demand, a lower heat load and a lower cooling requirement. By the cube law of Section 2.1.4 the trade is favorable: if an illustrative GPU would draw 400 W uncapped and all of its power scaled as the cube of frequency, a 350 W cap would require a clock reduction to $(350/400)^{1/3} = 0.956$ of the original, a performance loss of about 4.4 % for a power reduction of 12.5 %. A static share of the power makes the required frequency reduction somewhat larger, and a job limited by memory bandwidth rather than by the clock loses less, so the real trade depends on the workload. Across an illustrative cluster of 1000 such GPUs, the cap lowers the load by $1000 \times 50\,\text{W} = 50\,\text{kW}$, and the heat by the same amount.

The cap is a control variable. In this subsection the symbol u changes meaning: it denotes the control input, here the cap itself, $u = P_{\text{cap}}$, and no longer the utilization. The operator's choice of caps can be written as an optimization,

$$\begin{aligned} \min_{u} \quad & J(x, u) \\ \text{subject to} \quad & T_{\text{rack}} \leq T_{\max}, \\ & P_{\text{facility}} \leq P_{\text{facility}}^{\max}, \\ & \text{required computing demand served,} \end{aligned} \tag{2.12}$$

where x is the state of the system (the loads, the temperatures, the jobs in progress), J is a cost function that collects the energy cost and the cost of the lost or delayed computation, T_{rack} is a rack temperature with its limit $T_{\max}$, and $P_{\text{facility}}^{\max}$ is the limit on the facility's total power, for example the capacity of its grid connection. The formulation says that power caps are chosen to keep temperatures and total power within limits while still serving the computing that must be served, at least cost. It also shows that the cap is not only an electrical setting: it is an actuator for the thermal constraint, because lowering P_{IT} lowers $\dot{Q}_{\text{IT}}$ by equation (2.7). Control loops of this kind are the subject of Chapter 10. Outside this subsection and unless stated otherwise, u again denotes utilization.

2.4.4 Shifting work in time

Not all computing has to happen at once. Some work is immediate: web requests, financial transactions and interactive inference, in which a trained model answers a user, must be served within fractions of a second or the service fails. Other work is flexible: the training of machine learning models, batch analytics, backups and some scientific computing must be done, but not at a particular minute.

A flexible job is described by its energy and its window. Consider a job that needs $E_j = 100\,\text{kWh}$ and may run at any time between 00:00 and 06:00. Its power profile $P_j(t)$ must satisfy

$$\int_{00:00}^{06:00} P_j(t)\,\mathrm{d}t = E_j = 100\,\text{kWh}, \qquad 0 \leq P_j(t) \leq P_j^{\max}, \tag{2.13}$$

where $P_j^{\max}$ is the most power the hardware assigned to it can draw. The operator chooses when the load appears: spread evenly over the six hours it is $100/6 = 16.7\,\text{kW}$; concentrated between 02:00 and 04:00 it is 50 kW for two hours and zero otherwise; any other profile with the same area and within the bound is admissible. The energy is fixed; its timing is a decision. The

heat follows the same profile, so a job moved into the coolest hours of the night is also cooled by a plant operating at its most efficient. This makes computing a flexible energy resource, in the same sense as a deferrable load in a power system: a quantity of energy that must be delivered by a deadline, at a time the operator can choose to suit prices, grid conditions or cooling capacity.

2.4.5 Shifting work between sites

An operator with several data centers can also move work between them. Suppose data center DC1 is constrained, by a high outdoor temperature that reduces its cooling capacity or by a high electricity price, and DC2 is not. Moving work from DC1 to DC2 lowers the workload W_1 and raises W_2; through the chain of equation (2.8) it lowers the power P_1 and raises P_2. With illustrative numbers, moving work worth 2 MW of IT power from DC1, which during a heat wave draws 1.5 MW of facility power per megawatt of IT, to DC2, which draws 1.2, lowers DC1's facility load by $2 \times 1.5 = 3$ MW and raises DC2's by $2 \times 1.2 = 2.4$ MW; DC1's constraint is relieved and the total falls by 0.6 MW. The ratio of facility power to IT power used here is the power usage effectiveness of Chapter 11.

The move is bounded by the computing systems, not by the electrical ones: the data the work needs must be available at DC2 or moved there, the response time for users must remain acceptable from the new location, some data may be required by law to stay in a particular jurisdiction, and DC2 must have spare servers and spare cooling. Within those bounds, geographic shifting ties computing decisions to grid constraints, electricity prices and cooling conditions across a whole fleet of sites.

2.5 Variation in time and the coupled facility load

The IT load changes on every timescale from microseconds to seasons, and every change propagates into the facility's cooling and electrical systems. This section describes the timescales, shows why a forecast of IT load is useful to the plant, and writes the facility load as a sum whose cooling term depends on the IT term.

2.5.1 Timescales of variation

The IT load varies on four broad timescales, each with its own cause. From milliseconds to seconds, processor activity changes with every instruction stream, request and memory access. Over minutes, jobs begin and end, and GPU clusters ramp up or down as a training job starts, pauses or finishes. Over hours, interactive services follow the daily pattern of their users. Over days and seasons, the workload changes with the customers a facility serves, with holidays and with the growth of the installed hardware. Table 2.2 collects the four with what responds to each on the facility side.

Figure 2.5 shows an illustrative day of IT load for a facility of about 30 MW. The horizontal axis is the time of day in hours; the vertical axis is the IT power in megawatts. The solid thin line is the load sampled every two minutes; the dashed line is its daily-cycle component alone. Three of the four timescales are visible. The dashed curve is the hourly scale, a daily cycle between about 27 and 31 MW. The step at 10:00 is the minute scale at its largest: a GPU cluster ramps up and takes the load from about 30 to about 38 MW in thirty minutes, holds until 17:00 and then releases it in ten minutes; the smaller steps throughout the day are other jobs starting and ending. The jitter of the thin line stands for the fastest scale: each two-minute sample is an average over activity that changes far faster, and the sampling hides most of it.

Because the load has this structure, it can be forecast. The IT load is a time series, and a

Table 2.2: Timescales of IT load variation, their causes, and what the facility can do at each. The facility responses are indicative.

Timescale	Cause in the IT load	Facility response
milliseconds to seconds	processor activity, individual requests	absorbed by power electronics and by the thermal mass of servers and air; invisible to the plant
minutes	jobs start and end; GPU clusters ramp	fan speeds, valve positions and pump speeds adjust
hours	daily pattern of users and batch windows	chillers are staged on and off; supply temperatures are reset
days and seasons	customers, holidays, hardware growth	maintenance scheduling; capacity planning

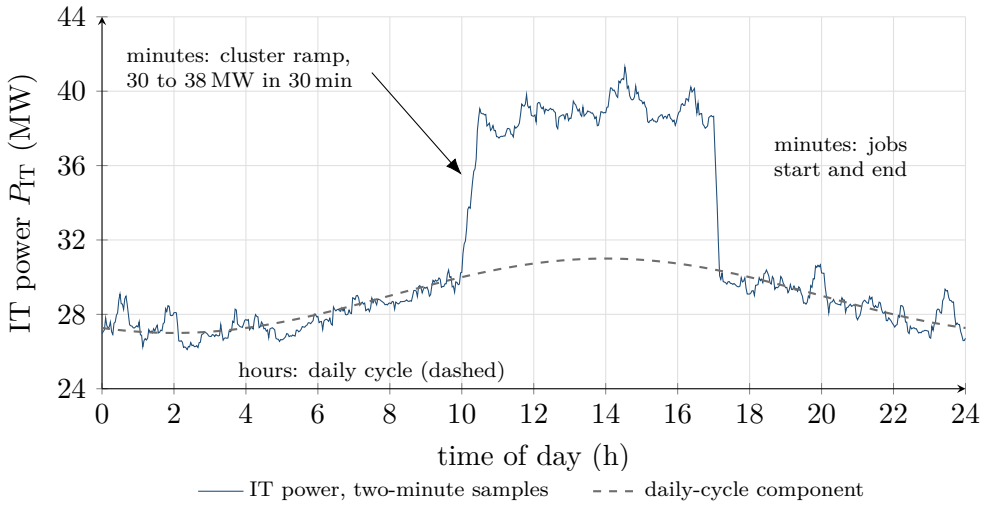


Figure 2.5: IT load over one illustrative day, sampled every two minutes, generated by `scripts/ch02_data.py` (seed 2). The load is the sum of a daily cycle $29+2\sin(2\pi(t-8)/24)$ MW (t in hours), an 8 MW cluster job ramped up over 10:00–10:30 and down over 17:00–17:10, smaller jobs of 0.2 to 0.6 MW lasting 5 to 40 minutes, and a fluctuation with standard deviation 0.1 MW standing in for sub-sample activity.

forecast of it h steps ahead takes the form

$$\hat{P}_{\mathrm{IT},t+h} = f(P_{\mathrm{IT},t},\ P_{\mathrm{IT},t-1},\ \ldots,\ \text{time of day and calendar},\ \text{workload information},\ \ldots), \tag{2.14}$$

a function of the recent values (which carry the persistence of the load), of time (which carries the daily and weekly cycles) and of what is known about the workload. The last argument has no counterpart in most electrical loads: the scheduler's queue lists jobs that have been submitted but not yet started, with their requested hardware, so part of the future IT load is announced before it appears.

2.5.2 Forecasting the IT load

A forecast is operationally useful because the cooling plant is slower than the load. If the forecast says that P_{IT} will rise from 30 MW to 38 MW in the next thirty minutes, as at 10:00 in Figure 2.5, the plant has time to prepare: start another chiller, raise the chilled-water flow, lower the supply temperatures to build margin, raise pump speeds. Starting a chiller and bringing it to stable operation takes minutes to tens of minutes; a plant that waits for the return water to warm before acting meets the new load late, with its temperatures above their setpoints while it catches up.

The forecast of power is a forecast of heat. By equation (2.7),

$$\dot{Q}_{\mathrm{IT}}(t) \approx P_{\mathrm{IT}}(t), \tag{2.15}$$

so a forecast of an 8 MW rise in IT power is a forecast of an 8 MW rise in the heat the plant must remove. With an illustrative chilled-water temperature rise of 6 K, the extra heat requires an extra water flow of

$$\dot{m}_w = \frac{\dot{Q}}{c_{p,w}\,\Delta T} = \frac{8000\,\mathrm{kW}}{4.18\,\mathrm{kJ/(kg\,K)} \times 6\,\mathrm{K}} \approx 320\,\mathrm{kg/s},$$

a figure the plant can schedule in advance if it has the forecast. One qualification applies: the heat appears in the servers at once, but it reaches the chilled water only after it has crossed the air paths and the coils, which takes from seconds to minutes; the dynamics of that delay are the subject of Chapter 8.

2.5.3 The facility load as a coupled load

The power a data center draws from the grid is more than its IT power. The facility load is the sum

$$P_{\mathrm{facility}} = P_{\mathrm{IT}} + P_{\mathrm{cool}} + P_{\mathrm{loss}} + P_{\mathrm{aux}}, \tag{2.16}$$

where P_{cool} is the power of the cooling system (chillers, pumps, fans, cooling towers), P_{loss} the losses of the electrical distribution (transformers, uninterruptible power supplies, cables) and P_{aux} the auxiliary loads (lighting, offices, security systems). The terms are not independent. The cooling power depends on the heat it removes, and therefore on the IT power:

$$P_{\mathrm{cool}} = f(P_{\mathrm{IT}},\ T_{\mathrm{amb}},\ \text{equipment},\ u), \tag{2.17}$$

where T_{amb} is the outdoor temperature, which sets how hard the plant must work to reject heat, the equipment is the installed plant with its efficiencies, and u here again denotes the control settings of the cooling plant (setpoints, speeds, which units are running), not utilization. The distribution losses also grow with the power carried, partly as its square through the resistive losses in transformers and cables. Only the auxiliary loads are roughly constant.

The consequence is that raising the IT demand raises the facility load by more than the server power. Every additional kilowatt in a server is a kilowatt of heat for the cooling plant, which spends electrical power to remove it, and a kilowatt that the electrical distribution carries with some loss. The electrical side of this coupling is the subject of Chapter 3, and the efficiency measure that summarizes it, the ratio $P_{\mathrm{facility}}/P_{\mathrm{IT}}$, of Chapter 11.

2.5.4 A facility example

Take an illustrative facility with $P_{\text{IT}} = 10\,\text{MW}$ and 2 MW of cooling and auxiliaries, so that $P_{\text{facility}} = 10 + 2 = 12\,\text{MW}$. The IT load rises to 12 MW, and the cooling and auxiliaries rise to 2.5 MW; the facility now draws

$$P_{\text{facility}} = 12 + 2.5 = 14.5\,\text{MW}.$$

A 2 MW increase in computing caused a 2.5 MW increase in facility demand.

Figure 2.6 draws the two states as stacked bars, with heights to scale in megawatts. The left bar is the facility before the change and the right bar after it, each split into IT power (lower segment) and cooling and auxiliaries (upper segment). The dotted lines carry the top of the IT segment and the top of the bar from the left bar across to the right one, and the middle column lists the increments. The IT segment grows by 2 MW, the cooling and auxiliary segment by 0.5 MW, and the whole bar by 2.5 MW.

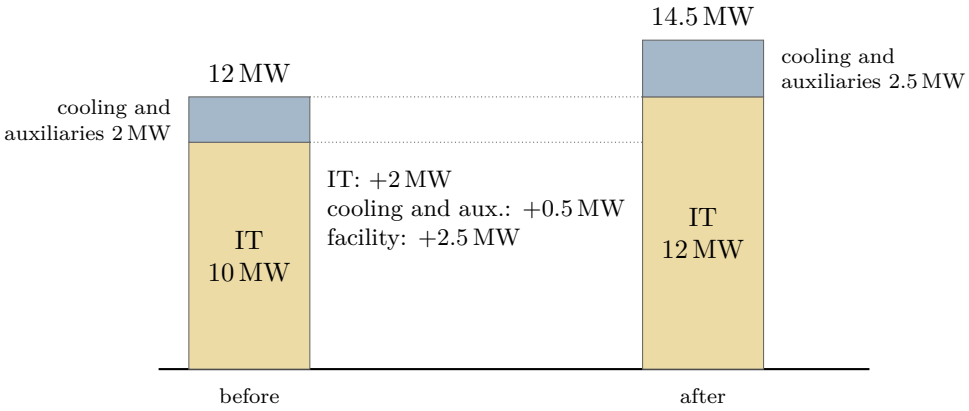


Figure 2.6: The coupled facility load, equation (2.16), illustrative. An increase of 2 MW in IT power raises the cooling and auxiliary load by 0.5 MW, so the facility load rises by 2.5 MW, from 12 to 14.5 MW.

The increment is larger in proportion than the average. Before the change the facility drew $12/10 = 1.20\,\text{MW}$ per megawatt of IT; after it, $14.5/12 = 1.21$; but the increment itself cost $2.5/2 = 1.25\,\text{MW}$ per megawatt of IT. The marginal cost of IT power exceeds the average when the cooling plant works harder per unit of heat at higher load, for instance because fan and pump power rise faster than the flow they deliver, or because an additional, less efficient chiller has to be started; the resistive part of the distribution losses also grows faster than the load. The difference matters for capacity planning. A facility whose grid connection is limited to 15 MW has, after the change, 0.5 MW of headroom, and that headroom supports not 0.5 MW of additional IT but, at the marginal rate of the example, $0.5/1.25 = 0.4\,\text{MW}$. A planner who used the average ratio would overestimate the IT that fits.

> **Principle 2.4.** *IT load is a coupled load. Each additional unit of IT power adds its own heat to the cooling plant and its own losses to the distribution, so the facility load rises by more than the IT load, and at the margin by more than the average ratio suggests.*

2.6 Telemetry, diagnosis and the state-space view

The preceding sections described the IT load as a physical and computational process. This section asks what can be observed of that process, how observations from different layers combine to explain an event, and how the whole is written as a dynamic system whose coupling runs in both directions.

2.6.1 Three layers of information

The electrical side of the IT load is measured directly: the rack power P_{rack} by the meters of the rack PDUs, the power of a PDU or a distribution panel P_{PDU} by its meter, and often the power of each server P_{server} by its baseboard management controller, the small independent computer on the server's main board that monitors its power, temperatures and fans. These are measurements. The rack heat $\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$ is a calculation from them. The true workload and the amount of useful computation are not electrical quantities at all; they come from software telemetry, the records that the operating systems, the schedulers and the applications keep about themselves.

There are three layers of information, which Figure 2.7 draws as three bands. The left column names the layer, the middle column lists its typical quantities and the right column the systems that record them; the arrows between the bands show the direction of cause. The top band is the business or workload telemetry: requests per second, jobs queued, response time (latency), service demand. The middle band is the computing telemetry: CPU and GPU utilization, memory use, job state (queued, running, finished, failed). The bottom band is the physical telemetry: voltage, current, power, temperature. The workload drives the computing activity and the computing activity drives the physical quantities, so the three are related; but they are not identical, because each relation in between depends on things the others do not see, the scheduler's policy, the workload type, the processor's frequency and the room's temperature. Each layer is typically kept by a different group in a different system with its own clock and sampling rate, which makes the three hard to align in time.

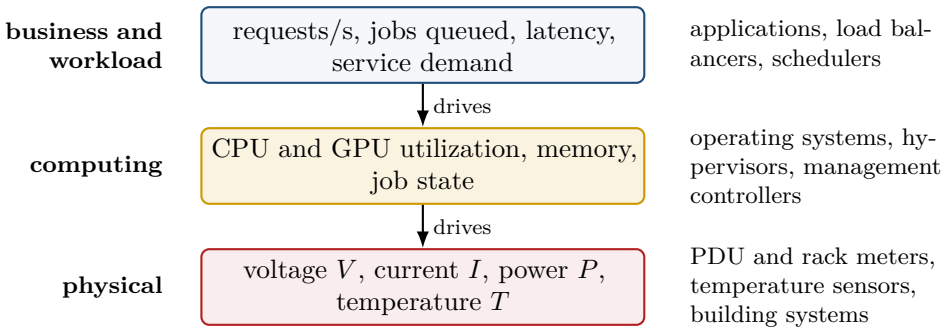


Figure 2.7: The three layers of information about an IT load. Left: the layer; middle: its typical quantities; right: the systems that record them. Each layer drives the one below it; the layers are related but not identical, and are usually recorded by different systems.

2.6.2 Diagnosis of a power drop

Suppose the measured power of a rack falls suddenly. Five hypotheses explain the observation: H_1, the workload dropped; H_2, servers were shut down; H_3, a power cap was activated; H_4, an electrical failure removed part of the rack's supply; H_5, the meter is in error. The electrical measurement alone cannot distinguish them, because each of the five produces the same symptom in the one quantity measured.

Table 2.3 lists what each hypothesis predicts for the other layers. The columns are the rack power, the GPU utilization u_{GPU}, the number of active servers N_{active}, the rack temperature T_{rack}, and other evidence that separates the hypothesis from its neighbors. A workload drop lowers utilization on servers that remain active; a shutdown lowers the number of active servers while the remaining ones keep their utilization; a power cap leaves the servers active and busy, at lower power and lower throughput; an electrical failure removes servers abruptly without any scheduler action, or, on a rack whose servers have two power cords, moves load from one feed to

the other; a meter error changes one reading and nothing else, so the rack temperature does not fall. Once the electrical measurement is combined with u_{GPU}, N_{active} and T_{rack}, the inference is much stronger: the hypotheses that agreed on the power disagree on the other quantities.

Table 2.3: Five hypotheses for a sudden drop in measured rack power and what each predicts in the other layers. Arrows mark the predicted direction of change; "same" marks no change.

Hypothesis	Rack power	u_{GPU}	N_{active}	T_{rack}	Separating evidence
H_1 workload dropped	↓	↓	same, then ↓	↓ slowly	fewer requests or queued jobs in the workload layer
H_2 servers shut down	↓	same on the rest	↓	↓ slowly	shutdown recorded by the scheduler or the operators
H_3 power cap activated	↓, then flat at the cap	same, high	same	↓ slowly	lower clock frequency; lower throughput; cap event logged
H_4 electrical failure	↓ abruptly	lost with the servers	↓ abruptly	↓ slowly	no scheduler action; alarms upstream; second feed of dual-corded servers rises
H_5 meter error	↓ on one meter	same	same	same	upstream meter and server readings unchanged; facility balance no longer closes

The result of such a combination is not a single verdict but a probability for each hypothesis given all the measurements. Evidence from several domains, electrical, computing and thermal, is what moves those probabilities apart, and when two hypotheses remain close, the useful next measurement is the one on which they disagree most: the clock frequency separates H_3 from H_1, an upstream meter separates H_5 from all the others. Diagnosis of this kind is developed in a later chapter; here the point is that it requires the three layers of Figure 2.7 together.

Principle 2.5. *One electrical measurement cannot explain a change in IT load. The explanation comes from combining physical, computing and workload evidence, because hypotheses that predict the same power predict different utilization, server counts and temperatures.*

2.6.3 The state-space view

The quantities of this chapter can be organized as a dynamic system. In this subsection u denotes the vector of control inputs, as in the notation chapter, while utilizations appear as components of the state with subscripts, u_{CPU} and u_{GPU}. The state collects the quantities that describe the IT load at an instant,

$$x = \left[N_{\text{active}},\ u_{\text{CPU}},\ u_{\text{GPU}},\ P_{\text{IT}},\ T_{\text{rack}},\ \dots\right]^{\mathsf{T}};$$

the control inputs are the decisions that act on it,

$$u = \left[P_{\text{cap}},\ \text{workload assignment},\ \text{server on/off state},\ \dots\right]^{\mathsf{T}};$$

and the disturbances are the external drivers that the operator does not choose,

$$d = \left[\text{incoming workload},\ T_{\text{amb}},\ \text{grid conditions},\ \dots\right]^{\mathsf{T}}.$$

The evolution of the state is written

$$\dot{x} = f(x, u, d, \theta), \tag{2.18}$$

where θ collects the parameters of the system, for example the idle and full-load power of each server type, the coefficients of the non-linear power curve, and the thermal capacitance of the racks.

The state is mixed. N_{active} is an integer that changes in jumps when the scheduler acts; the utilizations change almost instantly; the rack temperature changes slowly, through the thermal mass. P_{IT} is listed in the state as in many operational models, although by equation (2.2) it is largely an algebraic function of the server counts and utilizations, and could equally be treated as an output of the state rather than part of it. Equation (2.18) is the form into which the following chapters add the electrical distribution, the air, the water and the cooling plant as further states.

2.6.4 Two-way coupling

The chapter's chain runs in two directions. Forward, computational demand produces hardware activity, hardware activity produces IT power, IT power becomes heat, and heat becomes cooling demand. Backward, limits on power or cooling produce power caps or scheduling decisions, and those change computing performance. The three domains, computing, electrical and thermal, are coupled both ways: a cooling plant that cannot keep up forces caps that slow the computation, and a grid connection near its limit forces scheduling decisions that delay it.

Figure 2.8 draws both directions and a comparison. The upper part has one box per domain; the arrows above the boxes run forward, from computing to electrical to thermal, and the arrows below run backward, from thermal limits through electrical limits to computing performance. The lower part compares two chains. A distributed energy resource, such as a heat pump or an electric vehicle charger, connects customer demand to device power to grid behavior. A data center has the same structure with additional links: computing demand produces server power, server power becomes heat, and removing the heat costs cooling power, all of which reach the grid. The data center is therefore a multi-physics, controllable electrical load: its flexibility is set by computing decisions, and its total power by the thermal consequences of those decisions.

2.7 What to retain

Definition 2.1 (IT load)**.** IT load is the electrical power of the servers, storage and network equipment, set by computing demand through the scheduler's decisions and the physics of the processors, with an idle offset that makes it affine rather than proportional in utilization. Nearly all of it becomes heat, so it is at once an electrical load, a heat source and a controllable lever, coupled in both directions to the facility's power and cooling.

Table 2.4 lists the chapter's objects and the subsection that defines each.

2.8 Exercises

Exercise 2.1. A server has $P_{\text{idle}} = 150\,\text{W}$ and $P^{\max} = 500\,\text{W}$. (a) With the linear model, compute its power at utilization $u = 0.3$ and at $u = 0.8$. (b) The server runs 8 hours a day at $u = 0.8$ and 16 hours at $u = 0.3$; compute its daily energy and the fraction of it that is idle power. (c) A server of the same $P^{\max}$ whose power were proportional to utilization would draw how much energy over the same day?

Table 2.4: The objects of the IT load model and where each is defined.

Object	Expression	Role	Defined in
Server power	equation (2.1)	sum of component powers	2.1.1
Linear power model	$P_{\text{server}} = P_{\text{idle}} + u(P^{\max} - P_{\text{idle}})$	power from utilization, with idle offset	2.1.2
Processor power	$P_{\text{dyn}} \propto C_{\text{eff}} V^2 f_{\text{clk}}$, plus P_{static}	why frequency and voltage matter	2.1.4
Rack power and heat	$\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$	electrical-to-thermal link	2.2.1
Rack density	kW per rack, Table 2.1	sets the cooling architecture	2.2.2
Demand-to-heat chain	equation (2.8)	decision, physics, conservation	2.3.1
Arrival-rate model	$N_{\text{required}} \approx \lambda/\mu$	demand to number of servers	2.3.3
Group power	equation (2.10)	placement changes idle power	2.4.1
Power cap	$P_{\text{server}} \le P_{\text{cap}}$; problem (2.12)	control of power and heat	2.4.3
Deferrable job	equation (2.13)	fixed energy, chosen timing	2.4.4
IT load forecast	equation (2.14)	forecast of power and of heat	2.5.1
Facility load	equations (2.16)–(2.17)	IT load plus coupled terms	2.5.3
Information layers	Figure 2.7	physical, computing, workload	2.6.1
Hypotheses H_1–H_5	Table 2.3	explanations of a power drop	2.6.2
State-space model	$\dot{x} = f(x, u, d, \theta)$	states, controls, disturbances	2.6.3

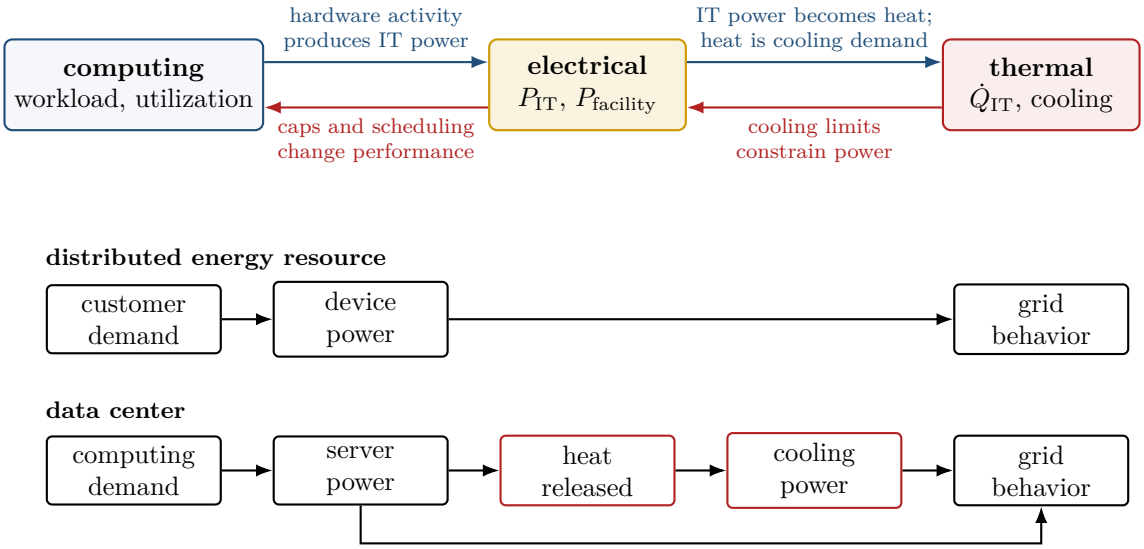


Figure 2.8: Two-way coupling of the computing, electrical and thermal domains (upper part): forward arrows carry demand to heat, backward arrows carry limits back to performance. Lower part: the chain of a distributed energy resource and the longer chain of a data center, whose server power reaches the grid both directly and through heat and the cooling power that removes it (red-outlined boxes).

Exercise 2.2. A service receives $\lambda = 8000$ requests/s, and each server handles $\mu = 400$ requests/s at full utilization. The servers have $P_{\text{idle}} = 180\,\text{W}$ and $P^{\max} = 450\,\text{W}$. (a) Compute the minimum number of servers and their total power at $u = 1$. (b) The operator instead runs 40 servers. Compute their utilization and total power. (c) Using equation (2.10), compute the saving of consolidating from 40 servers to 25 if the 15 freed servers are switched off, and the saving if they are left on and idle.

Exercise 2.3. A facility draws $P_{\text{IT}} = 20\,\text{MW}$ and 5 MW of cooling, distribution losses and auxiliaries, and its grid connection is limited to 30 MW. (a) The marginal cooling, loss and auxiliary load is 0.3 MW per megawatt of additional IT. How much IT power can be added before the connection limit is reached? (b) How much would a planner who used the average ratio $P_{\text{facility}}/P_{\text{IT}}$ predict, and by how much would that overestimate the IT that fits? (c) A forecast predicts that the IT load will rise by 3 MW in 30 minutes. With a chilled-water temperature rise of 5 K and $c_{p,w} = 4.18\,\text{kJ/(kg K)}$, compute the additional chilled-water flow the plant must provide.

Exercise 2.4. The measured power of a GPU rack falls from 40 kW to 38 kW within a minute. (a) Over the next hour, the GPU utilization stays at 92 %, the number of active servers is unchanged and the rack temperature falls slowly. Which hypothesis of Table 2.3 is most consistent, and which single additional reading would confirm it? (b) In a second incident, the rack meter shows the same drop but the utilization, the server count and the rack temperature are all unchanged, and the upstream panel meter shows no change. Which hypothesis remains? (c) If the rack's GPUs were capped in case (a) from an uncapped 400 W each to a cap of 350 W, and the whole drop is due to the cap, how many GPUs does the rack hold, what share of the rack's original power did they draw, and what clock reduction does the cube law of Section 2.1.4 imply?

2.9 Connection to the TERRA framework

The IT load is where the computing system enters the physical model of a data center, and the chapter's objects map onto the framework's graph as follows.

- **The rack is a node in two networks.** In the electrical network the rack is a load node whose inflow P_{rack} is measured by its PDU; in the thermal network it is a source node whose heat $\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$, equation (2.7), enters the air. The relation $\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$ is the coupling between the two conservation networks, written once per rack.
- **Server power is a component relation with parameters.** Equation (2.2) is the constitutive relation of a server component, with parameters $\theta = (P_{\text{idle}}, P^{\max})$ and utilization as its driving input. The parameters can be inferred from paired power and utilization telemetry, and the gap between the linear model and Figure 2.2's non-linear curve is a model error to be carried as uncertainty.
- **The facility sum is a redundancy relation.** Equation (2.16) is a balance at the facility's point of supply: with P_{facility}, P_{IT} and P_{cool} metered, its residual tests the meters and the loss model, and a residual that opens on one rack's meter alone is the signature of H_5.
- **A power drop is a ranked posterior over hypotheses.** H_1 to H_5 are competing explanations of one symptom; the framework assigns each a posterior probability from electrical, computing and thermal evidence together, as in Table 2.3, and the next measurement is chosen where the leading hypotheses disagree most.
- **The computing layer enters as inputs and disturbances.** Workload is not a conserved physical quantity, so it is not a network of its own: incoming work is a disturbance d, and caps, placement and on/off decisions are controls u of equation (2.18). A forecast of P_{IT} drives the forward simulation of the thermal network, and the risk of the 10:00 ramp of Figure 2.5 is the probability that the available cooling falls below the required cooling during it.

In summary, the IT load is an affine function of utilization with an idle offset, aggregated by racks into densities that fix the cooling architecture, driven by computing demand through the scheduler, shaped by consolidation, capping and shifting, and coupled to the facility's cooling and electrical systems in both directions; its explanation requires physical, computing and workload evidence together. The next chapter follows the IT power upstream, through the electrical distribution that delivers it from the grid to the rack.

Chapter 3

Electrical power distribution: from the utility connection to the server

3.1 Continuity of power

Chapter 2 described what the IT equipment consumes. This chapter follows that power backward, from the server to the point where it enters the facility, and asks what the equipment in between is for. The question of the chapter is

how a data center delivers power to its IT equipment without interruption, even when a piece of equipment or the utility supply itself fails.

The answer is not a single device but an arrangement: several sources, several distribution paths, protection that isolates a fault without spreading it, and stored energy that covers the seconds in which nothing else can. The chapter takes these in the order the power meets them, then turns to the questions an operator asks of the arrangement as a whole: how much spare capacity it has, how its losses close the facility's power balance, and how likely it is that a failure reaches the servers.

Jargon: switchgear, selective protection, transformer, UPS, double conversion, PDU, busway, dual-corded, N / N+1 / 2N, contingency

Switchgear is the assembly of circuit breakers, disconnect switches, protection relays, busbars and meters that receives power and divides it among circuits. *Selective protection* (also selectivity or coordination) is the setting of breakers and relays so that a fault is cleared by the protective device closest to it, leaving every other circuit energized. A *transformer* changes the voltage of alternating current, here from the medium voltage of the utility service to the low voltage used inside the building. A *UPS* (uninterruptible power supply) is a power converter with stored energy, usually a battery, that keeps its output supplied when its input fails. *Double conversion* is the UPS design in which all power is converted from alternating to direct current and back again, so that the output is regenerated rather than passed through. A *PDU* (power distribution unit) distributes, protects and meters power to a group of racks; a *rack PDU* does the same inside one rack. A *busway* is a rigid, enclosed set of conductors run above the rack rows, with plug-in tap boxes for individual racks. A *dual-corded* server has two power supplies, each fed from a different path. N, $N{+}1$ and $2N$ are redundancy levels: N units are exactly enough to carry the load, $N{+}1$ adds one spare, and $2N$ provides two complete and separate sets of N. A *contingency* is a credible failure of one component or path, together with the operating state that follows it.

Key idea

A data center's electrical system is designed for continuity of power, not only for efficient delivery: multiple sources feed multiple distribution paths through protection that isolates faults, and stored energy bridges the gap before backup generation starts. Its operating question is not whether the present load is below capacity but whether the load can still be carried after a credible failure.

3.1.1 The path from the utility to the server

Electric power reaches a server through a sequence of stages, each of which changes the power in one respect: its voltage, its protection, its continuity or its division into smaller circuits. Followed from the outside in, the sequence is

utility grid → switchgear → transformer → UPS → PDU or busway → rack → server.

The utility delivers medium-voltage power to the site. The switchgear receives it, protects the circuits and divides it. Transformers lower the voltage to the level the building's equipment uses. The UPS makes the supply continuous. The PDU or busway divides the protected supply into circuits for individual racks, and the rack PDU divides it once more among the servers, switches and storage units in the rack.

Figure 3.1 draws this sequence. Each box is one stage, in the order the power meets it from left to right; the text above each box is the voltage level or the role of the stage, with illustrative values used throughout the chapter. The dashed box below the switchgear is the on-site backup generation, which joins the path at the switchgear when the utility fails. The figure is a single path, and that is its limitation: in a real facility every stage is duplicated or backed by a spare, and the duplicated paths are connected so that power can reach a server by more than one route. The topology of a data center's electrical system is therefore a network rather than a chain, and most of this chapter is about the consequences of that difference.

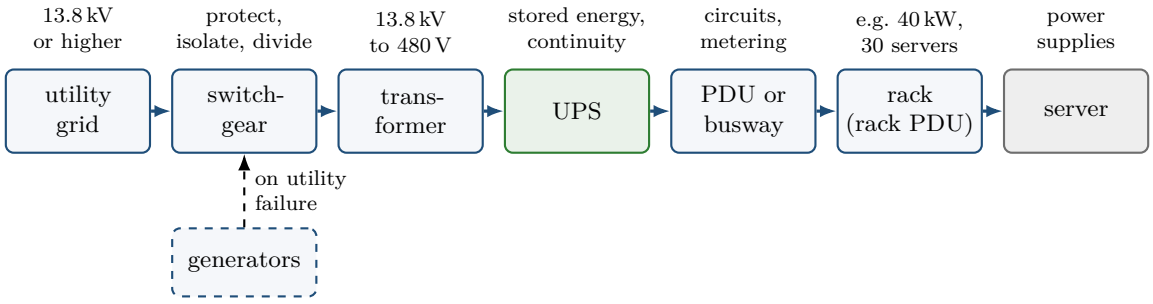


Figure 3.1: A single power path from the utility connection to a server. Voltages and the rack load are illustrative values used in the chapter. Real facilities duplicate every stage and interconnect the copies, so the electrical topology is a network, not this chain.

3.1.2 Why these systems differ

An office building and a data center both take power from the utility and distribute it to loads, but they are designed against different failures. For most loads a power interruption is an inconvenience: the lights go out and come back. For a server, an interruption of a fraction of a second is enough to stop it, and a stopped server loses the computations and transactions in progress, takes minutes to restart, and may leave the applications that depended on it in an inconsistent state. The design goal of a data center's electrical system is therefore *continuity*

of power, the absence of any interruption visible to the IT equipment, and only secondarily efficient delivery.

The list of things that threaten continuity is long, and each item is a credible event over the life of a facility:

- a utility outage, from a fault on the grid or a storm;
- a transformer failure;
- a breaker trip, intended or not;
- a UPS failure;
- a failure of a cable or of a busbar;
- maintenance, which takes equipment out of service on purpose;
- the start delay of the generators, during which neither the utility nor the generators supply the load;
- an overload, when a circuit carries more than its rating;
- a short circuit somewhere downstream.

The design principle that answers the list is that no single one of these events should stop the computing. The rest of the chapter shows how each stage of the path contributes to that principle and where it can fail to.

> **Principle 3.1.** *A data center's power system is designed so that no single credible failure, including the loss of the utility, interrupts the supply to the IT equipment.*

3.2 From the utility to the UPS

This section follows the power from the utility connection through the switchgear and the transformers to the UPS, and describes the generators that back up the utility. It establishes the two quantitative relations of the upstream path: why power is carried at high voltage, and how much energy the UPS must store.

3.2.1 The utility service

Power arrives at the site from the utility at medium or high voltage, for example 13.8 kV or higher. The connection, the set of utility lines, transformers and protection that link the site to the grid, sets the maximum power the facility can draw. A facility rated

$$P_{\text{facility,max}} = 100\,\text{MW} \tag{3.1}$$

needs an interconnection and internal electrical infrastructure able to deliver 100 MW, and from the utility's point of view it is a significant grid customer: comparable to a small town, and connected under a dedicated agreement that fixes its maximum demand.

The scale shows why voltage matters from the start. At unity power factor, 100 MW at 13.8 kV corresponds to a line current of about $100 \times 10^6/(\sqrt{3} \times 13\,800) \approx 4180$ A (the three-phase relation is derived in Section 3.2.3), which already requires several parallel feeders. Facilities of this size therefore often take service at transmission voltage and step down to medium voltage in their own substation; smaller ones take service at medium voltage directly. Either way, the utility connection is the first capacity limit in the path, and every later stage is sized below it.

3.2.2 Switchgear and selective protection

The switchgear is the first equipment inside the site boundary. It contains circuit breakers, which open a circuit automatically on a fault or on command; disconnect switches, which isolate equipment visibly for maintenance but do not interrupt load current; protection relays, which measure currents and voltages and decide when a breaker should open; busbars, the rigid conductors that carry power between the incoming supply and the outgoing circuits; and meters. Its jobs are to distribute power to the downstream circuits, to isolate faults, to protect equipment against currents it cannot survive, and to allow maintenance by isolating one part of the system while the rest stays energized.

The protection is arranged in layers. A short circuit on one downstream circuit draws a fault current that every breaker between the source and the fault sees. If the breaker nearest the source opened, the whole facility would go dark for one fault. Selective protection sets each device to act faster or at a lower current than the device upstream of it, so that the breaker closest to the fault opens first and the upstream breakers, which see the same current, stay closed. A fault on one rack circuit then de-energizes that circuit only, and the rest of the facility does not notice it. The settings that achieve this are a design result in their own right: they must be checked against the fault currents each configuration of the network can produce, and a change in topology, such as running from generators instead of the utility, changes those fault currents.

3.2.3 Transformers and the choice of voltage

Transformers lower the voltage from the utility level to the level the building equipment uses, for example from 13.8 kV to 480 V. A transformer is efficient but not lossless; its power balance is

$$P_{\text{in}} \approx P_{\text{out}} + P_{\text{loss}}, \tag{3.2}$$

with the loss made up of a nearly constant part in the magnetic core and a part in the windings that grows with the square of the current. For an illustrative 2 MW transformer at 99 % efficiency the loss is about 20 kW, all of it heat in the transformer's room or yard.

The reason to carry power at high voltage and lower it only near the loads is the relation between power, voltage and current. In a balanced three-phase system the real power is

$$P = \sqrt{3}\, V_{\text{LL}}\, I_{\text{L}} \cos\varphi, \tag{3.3}$$

where V_{LL} is the line-to-line voltage, I_{L} the line current, and $\cos\varphi$ the power factor, the cosine of the phase angle between voltage and current, equal to one for a purely resistive load. For a fixed power and power factor, the current is inversely proportional to the voltage. Each conductor of resistance R dissipates

$$P_{\text{loss}} = I^2 R, \tag{3.4}$$

so the conductor loss grows as the square of the current, and for a fixed delivered power as the inverse square of the voltage.

A worked example makes the scale concrete. Take an illustrative feeder carrying 2 MW at a power factor of 0.95. At 13.8 kV, equation (3.3) gives

$$I_{\text{L}} = \frac{2 \times 10^6}{\sqrt{3} \times 13\,800 \times 0.95} = \frac{2 \times 10^6}{22\,707} \approx 88\ \text{A},$$

and at 480 V

$$I_{\text{L}} = \frac{2 \times 10^6}{\sqrt{3} \times 480 \times 0.95} = \frac{2 \times 10^6}{789.8} \approx 2532\ \text{A}.$$

The voltage ratio is $13\,800/480 = 28.75$, so the current is 28.75 times larger at the low voltage, and through the same conductors the loss is $28.75^2 \approx 827$ times larger. With an illustrative

resistance of $0.01\,\Omega$ per conductor and three conductors, the loss is $3 \times 88^2 \times 0.01 \approx 233\,\mathrm{W}$ at $13.8\,\mathrm{kV}$ and $3 \times 2532^2 \times 0.01 \approx 192\,\mathrm{kW}$ at $480\,\mathrm{V}$, almost $10\,\%$ of the power carried. A low-voltage run of the same length would need conductors with hundreds of times the cross-section to keep the same loss. Power is therefore transported at high voltage and stepped down as close to the loads as practical, and low-voltage runs are kept short.

Principle 3.2. *For a fixed delivered power, current is inversely proportional to voltage and conductor loss to the square of current; power is carried at high voltage and transformed down near the loads.*

3.2.4 The uninterruptible power supply

The UPS is the stage that turns a supply that can fail into one that does not, over short times. It sits between the upstream supply and the IT load,

$$\text{grid} \rightarrow \text{UPS (with stored energy)} \rightarrow \text{IT load},$$

and its defining property is what happens when its input disappears. If the input power falls to zero, $P_{\mathrm{grid}} \rightarrow 0$, the UPS supplies the load from its stored energy, $P_{\mathrm{UPS}} \rightarrow P_{\mathrm{IT}}$, with essentially no interruption visible to the servers. In the double-conversion design of Section 3.2.6 there is not even a switching event: the load is always fed by the UPS output, and only the source of the energy behind that output changes.

Why batteries when generators exist. A facility with generators still needs the UPS battery, because a generator cannot take the load instantly. On loss of the utility, the generator controls must detect the outage, start the engine, bring it to rated speed, stabilize the output voltage and frequency, and connect the generator to the bus through a transfer switch. Generators for this service are commonly specified to be ready to accept load within about ten seconds; that interval, short as it is, is two orders of magnitude longer than what a server tolerates. The battery carries the load across it:

grid fails → battery and UPS carry the load → generator starts → generator assumes the load.

The battery is a bridge, not the long-duration source. It is sized in minutes, with a margin for a generator that is slow to start or must be started a second time, while the generators are sized to run for hours or days on stored fuel. Chapter 4 follows this transition in detail.

3.2.5 The energy the UPS must store

A UPS that must carry a load P for a time t needs at least the energy

$$E = P\,t. \tag{3.5}$$

For an illustrative critical load of $P = 2\,\mathrm{MW}$ and a required support time of $t = 10\,\mathrm{min} = 1/6\,\mathrm{h}$,

$$E = 2\,\mathrm{MW} \times \tfrac{1}{6}\,\mathrm{h} = 0.333\,\mathrm{MWh} = 333\,\mathrm{kWh}.$$

This is the ideal figure, the energy delivered to the load. The battery installed must be larger, for five reasons: the inverter that converts the battery's direct current to the load's alternating current has losses, so the battery must deliver more than the load receives; the battery itself has an efficiency below one, more so at high discharge rates; the battery ages, and its capacity at the end of its service life is lower than when new; the battery may not be discharged completely without damage, so only part of its nameplate energy is usable; and the design adds a safety margin. With illustrative values, an inverter efficiency of 0.95, a usable fraction of 0.8 of

nameplate energy, and an end-of-life capacity of 0.8 of the new value, the nameplate energy must be

$$E_{\text{nameplate}} = \frac{333\,\text{kWh}}{0.95 \times 0.8 \times 0.8} = \frac{333\,\text{kWh}}{0.608} \approx 548\,\text{kWh},$$

about 1.6 times the ideal figure, before any safety margin. The power rating matters as much as the energy: the battery must deliver the full 2 MW, divided by the inverter efficiency, from the first instant.

3.2.6 The double-conversion topology

The common UPS design for data centers is double conversion. The incoming alternating current (AC) is converted to direct current (DC) by a rectifier; the DC feeds a DC bus to which the battery is connected; and an inverter converts the DC back to AC for the load:

$$\text{AC} \rightarrow \text{DC (rectifier)} \rightarrow \text{DC bus with battery} \rightarrow \text{AC (inverter)}.$$

In normal operation the power flows through both converters and the rectifier also keeps the battery charged. On an outage the rectifier stops delivering power, the battery, already connected to the DC bus, supplies it instead, and the inverter continues without interruption.

Figure 3.2 draws the arrangement. The box on the left is the incoming AC supply, from the utility or, after transfer, from a generator. The square with a diagonal is the rectifier, with the AC symbol (∼) on its input corner and the DC symbol (=) on its output corner; the square on the right is the inverter, drawn the other way round. The thick line between them is the DC bus, and the battery symbol, stacked long and short plates, hangs from it. The blue arrow at the top is the normal flow of power; the red dashed arrow is the flow during an outage, from the battery through the inverter to the load.

The figure also shows why the design isolates the IT load from the quality of the incoming supply. Because the load is always fed by the inverter, which synthesizes its output waveform from the DC bus, a voltage sag (a short drop), a swell (a short rise), a deviation of the frequency or a short disturbance on the input does not reach the load: it is absorbed by the rectifier and the DC bus, with the battery making up any shortfall. The price is that all of the power passes through two conversions and pays the losses of both, which Section 3.5.2 quantifies. Most units also have a bypass path around the converters, used for maintenance or when the inverter fails, which passes the input directly to the load without this isolation.

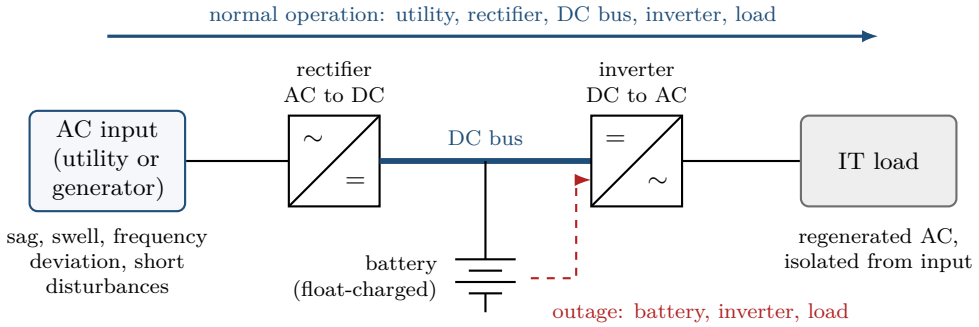


Figure 3.2: A double-conversion UPS. The rectifier converts the incoming AC to DC, the battery floats on the DC bus, and the inverter regenerates AC for the load. Input disturbances stop at the DC bus; on an outage the battery supplies the inverter with no switching event at the load.

3.2.7 Generators

Generators, usually diesel engines driving alternators, provide the longer-duration emergency power. They are connected to the facility through *automatic transfer switches* (ATS), switches

that disconnect the load from a failed source and connect it to another without operator action. The two operating states are

$$\text{normal: utility} \rightarrow \text{UPS} \rightarrow \text{IT load}; \qquad \text{failure: utility fails} \rightarrow \text{battery carries the load} \rightarrow \text{generator starts} \rightarrow \text{generator carries the load.}$$

When the utility returns and has been stable for a set time, the transfer is reversed and the generators are shut down after a cool-down run.

The generators must supply more than the IT load. The UPS protects the servers, but the servers keep producing heat, and the heat must keep being removed; chillers, pumps, air-handler fans, the control systems of the plant, and other critical auxiliaries such as fire protection and security must also run during an outage. Keeping the servers powered is not enough if cooling stops: within minutes the room temperature rises to the point where the servers protect themselves by shutting down, which is the interruption the UPS was installed to prevent. Chapter 5 quantifies how fast that happens. The generator plant is therefore sized for the facility load, P_{facility}, not for P_{IT} alone, and the cooling loads that are not on the UPS must restart on generator power quickly enough for the room to ride through the gap.

3.3 From the UPS to the server

Downstream of the UPS the power is continuous; the remaining stages divide it into circuits, carry it to the racks and deliver it to the servers' power supplies. This section describes those stages and the duplicated A and B paths that make each server independent of any one of them.

3.3.1 The power distribution unit

A power distribution unit receives the protected output of the UPS and delivers it to racks,

$$\text{UPS} \rightarrow \text{PDU} \rightarrow \text{rack}.$$

A PDU does four things. It steps the voltage down where the racks use a lower voltage than the UPS output, with a transformer inside the unit, for example from 480 V to 415 V three-phase with 240 V from line to neutral, or to the 208/120 V common in older North American facilities. It distributes the supply into individual branch circuits, one or more per rack. It measures the voltage, current and power of each circuit. And it protects each circuit with its own breaker, so that a fault in one rack opens one branch rather than the PDU.

Older designs place the PDUs as floor-mounted cabinets in the computer room, with cables run under a raised floor or overhead to each rack. Each new rack or change of rack load then needs a new cable from the PDU. Modern designs increasingly replace the floor-mounted PDU and its cables with busways or other overhead distribution, keeping the transformation and protection functions in fewer, larger units.

3.3.2 The busway

A busway is a rigid distribution backbone: a set of copper or aluminum bars in a metal enclosure, run along each row of racks above the cabinets, and fed at one end from the upstream panel or PDU. It acts as an electrical highway along the row. Each rack takes its power from a tap box, a plug-in unit with its own breaker and often its own meter, clipped onto the busway at the rack's position.

The arrangement has four advantages. Racks are easier to add, because a new rack needs a new tap box rather than a new cable run back to a PDU. Loads are easier to rearrange, because a tap box can be moved along the row. High currents are easier to carry, because the bars of a busway can be sized for several hundred amperes over the full length of the row. And circuits

are easier to instrument, because a meter in each tap box reports the power of each rack where it is drawn. The last point links electrical distribution to the rest of the facility's model: the power of each rack is the heat of each rack.

3.3.3 Rack power

Inside the rack, one or more rack PDUs, vertical strips of outlets with their own breakers and meters, distribute the supply to the equipment. An illustrative rack drawing

$$P_{\text{rack}} = 40\,\text{kW}$$

holds 30 servers, network switches, storage units and management hardware. If the switches, storage and management hardware take an illustrative 4 kW, the 30 servers share the remaining 36 kW, about 1.2 kW each. At 415 V three-phase and a power factor of 0.98, equation (3.3) gives a rack current of $40\,000/(\sqrt{3} \times 415 \times 0.98) \approx 57$ A per phase, a current that justifies the high-current busways of Section 3.3.2.

Rack monitoring reports, per rack PDU or per outlet, the voltage V, the current I, the real power P, the energy E accumulated over time, and the power factor. These readings serve two purposes. For electrical management they show how close each circuit is to its breaker rating and how the load is divided between the rack's power paths. For thermal modeling they are the best available estimate of the heat each rack releases into the room, since nearly all the electrical power of IT equipment leaves it as heat (Chapter 5). The book's distinction applies directly: the rack power is *measured*; the rack heat is *calculated* from it.

3.3.4 Dual power supplies and the A and B paths

Dual power supplies. Most servers in a data center have two power supply units, PSU_A and PSU_B, each connected by its own cord to a different power path. In normal operation the two supplies share the server's load, roughly half each; each is sized to carry the whole load alone. If path A fails, PSU_A loses its input and PSU_B takes the full load within the hold-up time of the server's internal capacitors, so the server keeps running. For the illustrative 1.2 kW server of Section 3.3.3, each supply normally delivers about 0.6 kW and must be able to deliver 1.2 kW.

The A and B paths. The dual supplies are useful only if the paths behind them are separate. A facility built for this duplicates the whole chain of Figure 3.1:

utility or generator A $\rightarrow$ UPS A $\rightarrow$ PDU A $\rightarrow$ PSU_A, utility or generator B $\rightarrow$ UPS B $\rightarrow$ PDU B $\rightarrow$ PSU_B.

The server survives the loss of either path, whatever stage of the path fails: a transformer, a UPS, a PDU, a busway or a single cord.

Figure 3.3 draws the two paths as a one-line diagram, the standard simplified drawing of a three-phase system in which one line stands for all three phase conductors. Path A runs down the left, path B down the right. At the top of each path, a utility feed (circle marked U) and a generator (circle marked G) enter an automatic transfer switch, which selects one of them. Below it, each path has a circuit breaker (small square), a transformer (two overlapping circles), a second breaker, the UPS, a third breaker and the PDU or busway. At the bottom the two paths meet only at the server, which has one supply on each path. The symbols are explained in the middle column. The drawing shows the property that the rest of the chapter examines: the two paths share no equipment above the server.

> **Principle 3.3.** *A dual-corded server on two separate paths survives the loss of any one stage of either path; the protection it gets is only as good as the separation of the paths behind its two cords.*

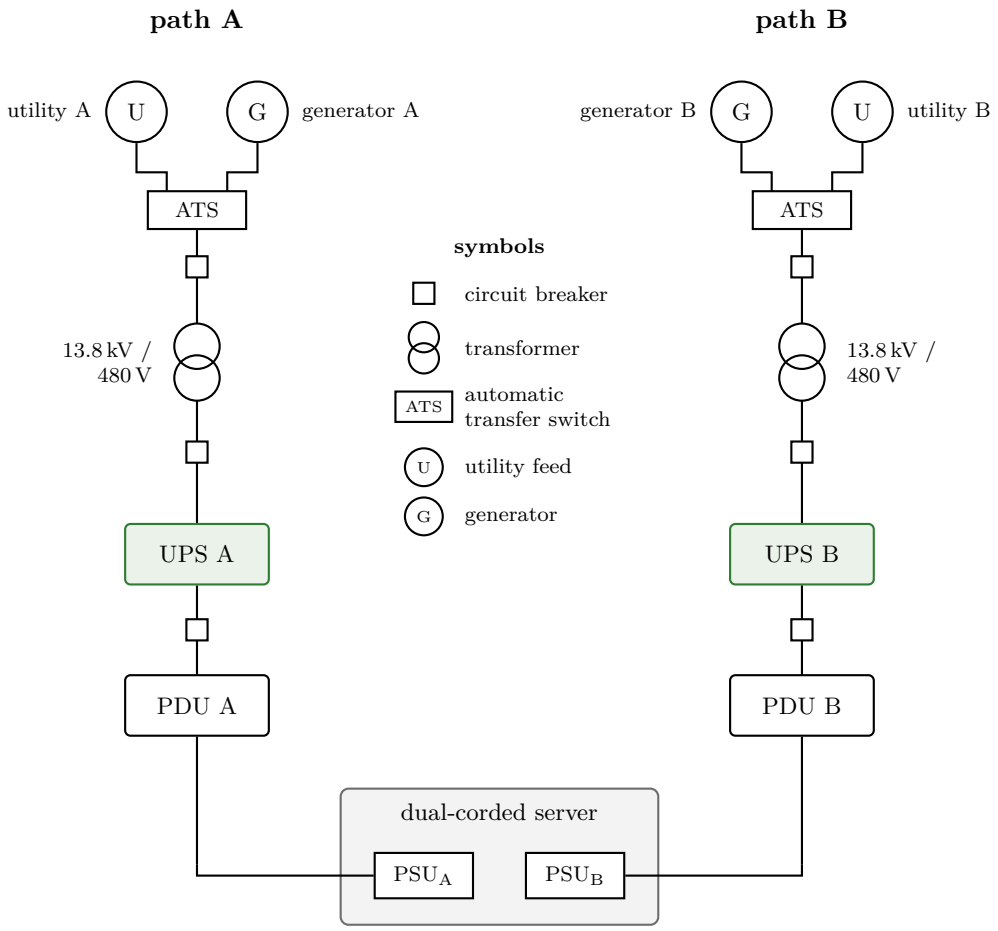


Figure 3.3: One-line diagram of two independent power paths to a dual-corded server. Each path has its own utility feed, generator, transfer switch, transformer, UPS and PDU; the paths meet only at the server, whose two power supplies each carry the full load if the other path fails.

3.4 Redundancy and contingency capacity

Duplicating equipment is the means by which the system survives a failure; this section names the levels of duplication, shows why duplication is not the same as independence, and states the capacity condition that each surviving path must meet after a failure.

3.4.1 N, N+1 and 2N

Redundancy levels are stated relative to N, the number of units exactly sufficient to carry the load. Consider an illustrative critical load that needs 4 MW of UPS capacity, built from UPS modules of 1 MW each.

- N: four modules, 4 MW installed. Every module is needed; each runs at 100 % of its rating at full load, and the failure of any one leaves 3 MW of capacity for a 4 MW load. There is no spare.
- N+1: five modules, 5 MW installed. Any one module may fail and the remaining four still carry 4 MW. At full load each module runs at $4/5 = 80\,\%$ of its rating. A module can also be taken out of service for maintenance, but during that time the system is reduced to N.
- $2N$: two independent systems of 4 MW each, 8 MW installed, the A and B systems of Section 3.3.4. Either system can carry the whole critical load. Each normally carries half, so each module runs at $2/4 = 50\,\%$ of its rating, and the facility survives the loss of an entire system, including its distribution, not only of one module.

The cost follows the installed capacity: $2N$ installs twice the capacity of N and 1.6 times that of N+1, and it is more robust in a specific sense. N+1 protects against the failure of a module inside one system but not against the failure of something the modules share, such as the common output bus or the switchgear that feeds them; $2N$ duplicates those shared parts as well.

Figure 3.4 draws the three levels. Each square is a 1 MW module. Filled squares are modules needed to carry the load; the dashed square in the N+1 row is the spare; the two groups in the $2N$ row are the separate A and B systems, each sufficient alone. The text on the right of each row gives the installed capacity, the loading of each module at full load, and the failure each level tolerates.

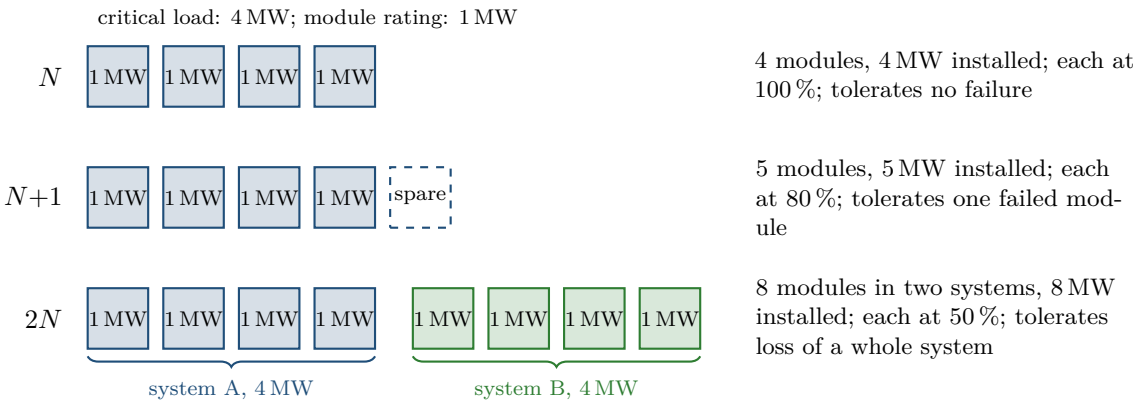


Figure 3.4: Redundancy levels for an illustrative 4 MW critical load built from 1 MW UPS modules. N has no spare, N+1 one spare module, and $2N$ two complete systems, either of which carries the whole load.

3.4.2 Redundancy is not independence

A $2N$ system is designed so that one path can fail without affecting the other. That is a statement about the equipment drawn on the one-line diagram, and it can be true of the diagram and false of the facility. Two nominally independent paths can depend on something the diagram does not show:

- the same upstream utility substation, so that one substation fault removes both utility feeds;
- the same cooling system, so that a cooling failure overheats both UPS rooms;
- the same control network, so that one control fault commands both systems wrongly;
- the same fuel supply, so that contaminated or exhausted fuel stops both sets of generators;
- the same room, so that a fire, a flood or a fire-suppression discharge reaches both.

A single event of any of these kinds, a *common-cause* failure, can take out both paths at once. The redundancy is real in the sense that two copies exist; it is not independence, because the copies do not fail independently.

The distinction has a quantitative consequence, developed in Section 3.6.3: the probability that both paths fail together is dominated by the common causes, not by the coincidence of two independent failures. Good designs physically separate the A and B paths into different rooms and fire zones, feed them from different utility substations where the grid allows it, and give them separate controls; the remaining shared dependencies are the ones an assessment of the facility has to find.

> **Principle 3.4.** *Redundancy is not independence: two paths that share a substation, a cooling system, a control network, a fuel supply or a room can be lost together to one event.*

3.4.3 Load sharing between paths

In a $2N$ arrangement with dual-corded servers, both paths carry load in normal operation. For an illustrative 10 MW critical load,

$$P_{\mathrm{A}} = 5\,\mathrm{MW}, \qquad P_{\mathrm{B}} = 5\,\mathrm{MW}.$$

If path A fails, the servers' A supplies lose their input and the B supplies take the full load, so path B must carry 10 MW. Its capacity must therefore satisfy

$$C_{\mathrm{B}} \geq 10\,\mathrm{MW}, \tag{3.6}$$

where C_{B} is the power rating of path B in megawatts (the symbol denotes a capacity here, not a thermal capacitance). The same holds for path A with the roles exchanged. The consequence for normal operation is that neither path of a $2N$ system may be loaded above half its rating; a path loaded to 60 % in normal operation would be asked for 120 % after the loss of its partner.

The requirement is the one that governs contingency analysis of a transmission system, in which every line must remain within its rating not only in the present state but after the outage of any other single line or transformer, the "$N-1$" criterion. A data center's paths are lines in this sense, and the failure of one is a contingency whose post-failure state must be feasible. The difference is that the consequence of an infeasible post-contingency state in a data center is immediate: the surviving path's breakers or UPS overload protection trip, and the load that both paths were protecting is lost.

3.4.4 Capacity margin before and after a contingency

The capacity margin of a path is its rating minus its load,

$$M = C - P. \tag{3.7}$$

An illustrative UPS path rated $C = 10\,\text{MW}$ carrying $P = 7\,\text{MW}$ has a margin of $M = 3\,\text{MW}$, and judged by its present state it is healthy: it runs at 70 % of its rating. Suppose now that a parallel UPS fails and 4 MW of its load transfers to this path. The post-contingency load is

$$P' = 7\,\text{MW} + 4\,\text{MW} = 11\,\text{MW} > C = 10\,\text{MW},$$

and the post-contingency margin is $M' = 10 - 11 = -1\,\text{MW}$: the path is overloaded by 10 %. The normal state satisfies $7 < 10$; the contingency state violates $11 > 10$. A margin of 3 MW is not enough for a transfer of 4 MW, and the present loading of 70 % says nothing about that. This is the logic of $N - 1$ contingency analysis applied to one path: the relevant margin is the one that remains after the largest credible transfer, not the one shown on the present loading display.

Figure 3.5 plots both examples. The vertical axis is the load on the surviving path in megawatts; the dashed red line is the 10 MW rating. The left pair of bars is the load-sharing example of Section 3.4.3: path B at 5 MW in normal operation (blue) and at 10 MW after the loss of path A (red). The right pair is the transfer example: 7 MW in normal operation and 11 MW after the 4 MW transfer. The label above each bar gives its load and margin. The first example reaches the rating exactly after the contingency, with no margin left for metering error or for growth of the load; the second exceeds it, although its normal-state margin of 3 MW is larger in absolute terms than any margin in the first example after the contingency.

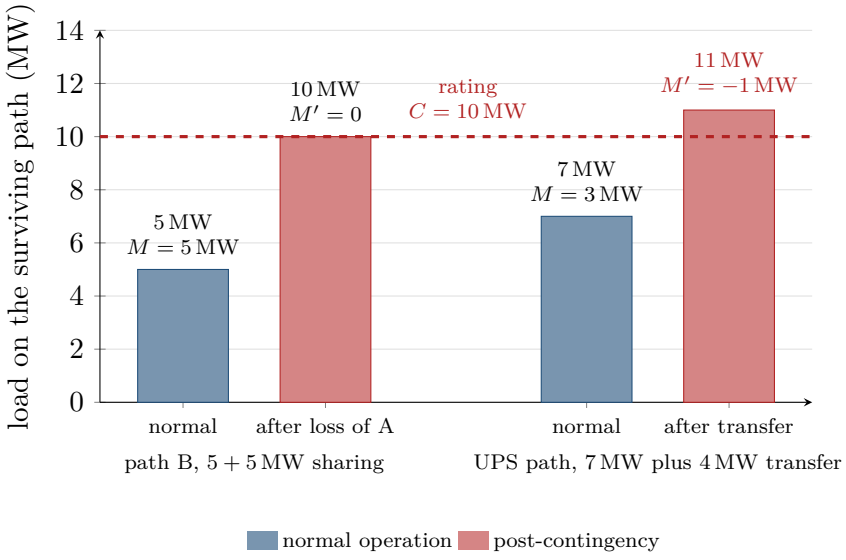


Figure 3.5: Path loading before and after a contingency against a 10 MW rating, illustrative. Left: two paths sharing 10 MW equally; after the loss of A, path B is exactly at its rating. Right: a path at 7 MW receives 4 MW from a failed parallel UPS and is overloaded by 1 MW. Margins follow $M = C - P$ (equation (3.7)).

Principle 3.5. *The margin that matters is the post-contingency margin, $C - P'$, after the largest credible transfer of load, not the margin of the present state.*

3.5 Losses and the facility power balance

The electrical system consumes part of the power it delivers, and the loads it serves are not only the IT equipment. This section links the electrical load to the cooling load, quantifies the conversion losses, writes the facility power balance, and treats a balance that does not close as an inference problem.

3.5.1 Electrical load and cooling are linked

The power the utility delivers is not all IT power. At the level of the whole facility,

$$P_{\text{facility}} = P_{\text{IT}} + P_{\text{cool}} + P_{\text{other}}, \tag{3.8}$$

where P_{cool} is the electrical power of the cooling system and P_{other} collects the conversion losses, lighting, offices and other auxiliaries. In an illustrative facility, an IT load of $P_{\text{IT}} = 80\,\text{MW}$ may correspond to $P_{\text{facility}} = 95\,\text{MW}$: 15 MW of cooling, losses and auxiliaries for 80 MW of computing, a ratio of $95/80 \approx 1.19$, the quantity Chapter 11 defines as the power usage effectiveness.

The consequence for capacity is direct. For the 100 MW connection of equation (3.1), judged from the server load alone the facility would seem to have $100 - 80 = 20\,\text{MW}$ of headroom; judged from the facility load it has $100 - 95 = 5\,\text{MW}$. Moreover the cooling term grows with the IT load, since every additional megawatt of IT power is a megawatt of heat to remove, and it grows with the outdoor temperature, since the cooling plant works harder on a hot day. A facility's electrical capacity cannot be judged from server load alone: the IT load, the cooling load and the losses are coupled, and the binding moment is usually a hot afternoon at high IT load.

3.5.2 Conversion losses

Every conversion stage dissipates part of the power that passes through it. For the UPS, with efficiency η_{UPS} defined as output over input,

$$P_{\text{in}} = \frac{P_{\text{out}}}{\eta_{\text{UPS}}}, \qquad P_{\text{loss,UPS}} = P_{\text{in}} - P_{\text{out}} = P_{\text{out}}\left(\frac{1}{\eta_{\text{UPS}}} - 1\right). \tag{3.9}$$

With an illustrative $\eta_{\text{UPS}} = 0.97$ and an output of 10 MW,

$$P_{\text{in}} = \frac{10\,\text{MW}}{0.97} = 10.31\,\text{MW}, \qquad P_{\text{loss,UPS}} = 10.31 - 10 = 0.31\,\text{MW}.$$

Over a year of continuous operation that is $0.31\,\text{MW} \times 8760\,\text{h} \approx 2700\,\text{MWh}$ of energy that never reaches the IT equipment.

Almost all of the loss becomes heat in the UPS room, and that heat has to be removed by the cooling system, which consumes electrical power of its own to do so. Electrical inefficiency therefore creates cooling demand: a loss is counted twice in the facility's consumption, once as the loss and once as part of the cooling power needed to remove it. Losses also depend on the operating point. The efficiency of a UPS typically falls at low load, where its fixed losses are a larger fraction of a smaller throughput, and a $2N$ system whose paths each run below 50 % of their rating by design (Section 3.4.3) spends its life in that region. The transformers, PDU transformers and the server power supplies add their own conversion losses in the same way, each with its own efficiency curve.

3.5.3 The facility power balance

Conservation of energy at the facility's point of connection gives a balance that every term of the electrical system must satisfy. Writing the cooling power by its parts where these are metered separately,

$$P_{\text{grid}} = P_{\text{IT}} + P_{\text{chillers}} + P_{\text{pumps}} + P_{\text{fans}} + P_{\text{loss,UPS}} + P_{\text{loss,tr}} + P_{\text{lighting}} + \dots, \tag{3.10}$$

where P_{grid} is the power measured at the utility meter (equal to P_{facility} while the facility runs on the utility), the chiller, pump and fan terms together make up P_{cool}, and $P_{\text{loss,tr}}$ is the transformer loss. The terms are of three kinds. P_{grid}, P_{IT} (at the UPS outputs or the PDUs) and the chiller, pump and fan powers are usually *measured* by meters. The UPS and transformer losses are usually *calculated*, either as the difference between input and output meters or from an efficiency curve and the measured throughput. And the trailing dots are the loads nobody listed.

If all the major terms are metered, the balance should close: the *residual*, the measured supply minus the sum of the measured and calculated loads,

$$r = P_{\text{grid}} - \sum_{k} P_k, \tag{3.11}$$

should be zero up to the meters' accuracy. If it is not, something is unmetered, mismeasured, mislabeled or incorrectly modeled. The balance is therefore not only bookkeeping: it is a test that the facility's picture of its own electrical system is complete and consistent.

3.5.4 A balance that does not close

Consider an illustrative reading in which the utility meter shows $P_{\text{grid}} = 12.0\,\text{MW}$ and the measured and calculated loads sum to 11.1 MW. The residual is

$$r = 12.0 - 11.1 = 0.9\,\text{MW},$$

7.5 % of the supply, unaccounted for. Figure 3.6 draws the balance as a node. The circle is the facility's main bus, the node at which equation (3.10) is written. The arrow entering from the left is the measured utility supply. The arrows leaving to the right go to the loads, each box giving an illustrative value and the gray label beside it saying whether that value is measured or calculated; the brace marks their sum. The IT load of 8.0 MW is metered at the UPS outputs, and its UPS loss of 0.25 MW is calculated from equation (3.9) with $\eta_{\text{UPS}} = 0.97$ ($8.0/0.97 - 8.0 = 0.247\,\text{MW}$). The dashed red arrow is the 0.9 MW that the listed loads do not account for.

The first question is whether 0.9 MW could be measurement noise. With illustrative accuracies of 0.5 % for the utility meter and 1 % for each submeter, and independent errors, the standard deviation of the residual is dominated by the three largest meters,

$$\begin{aligned}\sigma_r &\approx \sqrt{(0.005 \times 12.0)^2 + (0.01 \times 8.0)^2 + (0.01 \times 2.0)^2} \\ &= \sqrt{0.0036 + 0.0064 + 0.0004} \approx 0.10\,\text{MW},\end{aligned}$$

the smaller meters adding little. A residual of 0.9 MW is about nine standard deviations, not noise. Something is missing or wrong, and there are at least five candidate explanations:

- H_1, *meter bias*: the utility meter reads high or a submeter reads low, for example through a wrong current-transformer ratio in its configuration;
- H_2, *unmetered auxiliary load*: a panel feeding heaters, office air conditioning, battery-room ventilation or any other load that no meter covers;

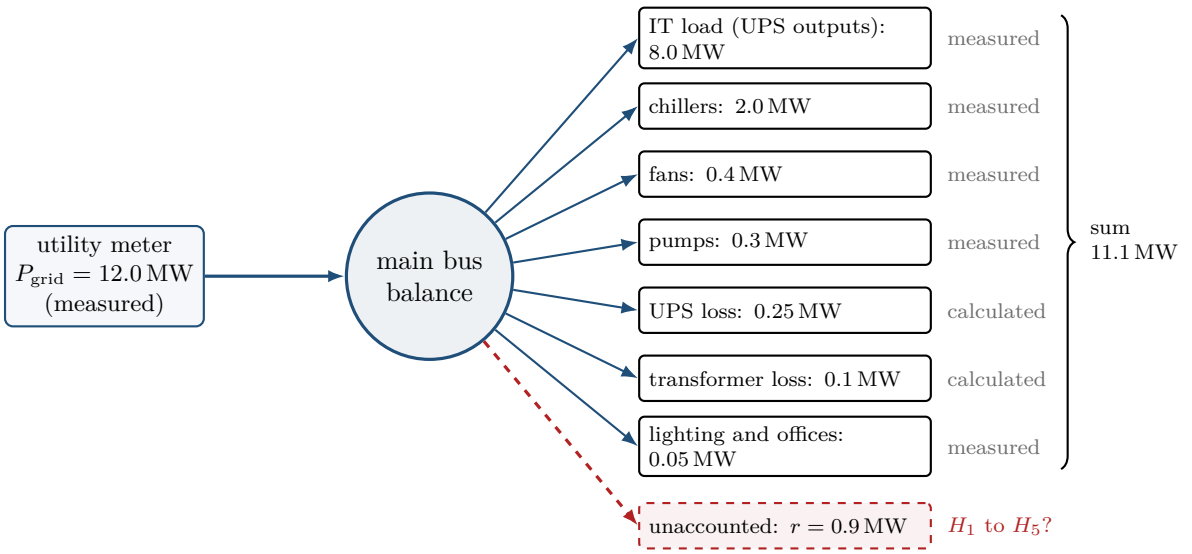


Figure 3.6: The facility power balance as a node, illustrative values. The measured supply of 12.0 MW leaves through seven measured or calculated loads summing to 11.1 MW; the residual of equation (3.11), 0.9 MW, is unaccounted for.

- H_3, *UPS loss underestimated*: the calculated 0.25 MW assumes an efficiency the units do not reach at their actual load;
- H_4, *cooling load missing*: part of the cooling plant, such as a cooling tower fan bank or a set of computer-room air handlers, is not in the metered sum;
- H_5, *topology or model error*: a load is assigned to the wrong meter or panel, counted twice, or not counted at all because the recorded wiring differs from the physical wiring.

A threshold alarm on the residual reports only that the power balance does not close. A diagnosis asks which explanation the evidence supports, and states the answer as a probability of each hypothesis given the measurements y. By Bayes' rule,

$$\Pr(H_i \mid y) = \frac{p(y \mid H_i)\,\Pr(H_i)}{\sum_j p(y \mid H_j)\,\Pr(H_j)}, \tag{3.12}$$

where $\Pr(H_i)$ is the prior probability of each hypothesis, from experience with similar facilities, and $p(y \mid H_i)$ is how well the hypothesis predicts the measurements. A single reading of 0.9 MW predicts nothing to separate them; each hypothesis can be sized to produce exactly that residual. The residual's behavior over time does separate them, because each predicts a different pattern. Table 3.1 lists the pattern each predicts and the measurement that would test it. A residual that follows the outdoor temperature points to H_4; one that scales with the IT load points to H_3 or to a gain error in the IT meters under H_1; one that is constant day and night points to H_2; and one that appeared as a step on the date of a reconfiguration points to H_5. The hypotheses are not exclusive: a 0.9 MW residual can be 0.5 MW of one cause and 0.4 MW of another, which the same reasoning handles by treating combinations as hypotheses or the size of each contribution as an unknown. The systematic treatment of such diagnoses is the subject of a later chapter.

Principle 3.6. *A facility power balance that does not close is evidence, not only an alarm: each candidate explanation predicts its own pattern for the residual, and the explanation is chosen by comparing those predictions with the measurements.*

Table 3.1: Explanations of the illustrative 0.9 MW residual of the facility power balance, the pattern each predicts for the residual over time, and a measurement that would test it.

Hypothesis	Predicted pattern of the residual	Measurement that tests it
H_1 meter bias	constant (offset error) or proportional to the biased meter's reading (gain error)	a temporary reference meter beside the suspect meter; a check of configured current-transformer ratios
H_2 unmetered auxiliary load	roughly constant, or following its own schedule (offices, heaters)	a survey of panels with a portable clamp meter; the panel schedule against the meter list
H_3 UPS loss underestimated	grows with the IT load; relatively larger at low load	simultaneous input and output metering of the UPS modules
H_4 cooling load missing	follows outdoor temperature and cooling demand	correlation of the residual with outdoor temperature; metering of the cooling panels
H_5 topology or model error	a step at the date of a reconfiguration; paired residuals of opposite sign at two downstream meters	inspection of breaker and switch positions; per-panel sums against the recorded wiring

3.6 Topology, state and reliability

The preceding sections assumed that the facility's record of its own wiring is correct and that every path fails independently. This section relaxes both assumptions, distinguishes the electrical state that is measured from the state that must be inferred, and assembles the chapter's material into a single model of the facility as a small power system.

3.6.1 Topology matters

Every calculation in this chapter uses the topology, the record of what is connected to what. Consider a rack physically connected to PDU 4 but recorded in the facility's database as connected to PDU 3. Every measurement can be accurate: the PDU meters, the rack PDU meters and the utility meter all read correctly. Yet every conclusion that combines the measurements through the topology is wrong. With the illustrative 40 kW rack, the sum of the racks recorded on PDU 3 exceeds PDU 3's meter by 40 kW, and PDU 4's meter exceeds the sum of its recorded racks by 40 kW: a pair of residuals of equal size and opposite sign, the signature of hypothesis H_5 in Table 3.1. Worse, a contingency assessment based on the record computes the wrong post-failure loads: the failure of PDU 4's upstream path is assessed without the rack, and the rack's exposure to it is missed.

Uncertainty in a data center's electrical model therefore lives in three places: in the measurements, which have errors and biases; in the parameters, such as efficiencies and ratings, which are known only approximately; and in the topology, which is a recorded belief about the wiring rather than an observation of it. The third is easy to overlook because the record looks like a fact. In a facility where racks are added, moved and rewired continually, it is an estimate that drifts.

3.6.2 Measured and inferred electrical state

The meters of the electrical system measure, at their own locations, the voltage V_i, current I_i and power P_i. Much of what an operator needs to know is not among these readings. Circuit loading is calculated from a measured current and a recorded rating where the circuit has its own meter, but must be inferred where it is assembled from downstream meters through the topology. The actual path by which each load is supplied, the state of transfer switches that do not report their position or report it unreliably, whether a UPS module or a generator is actually available to take load, and the effective remaining capacity after a credible failure are all inferred: they are conclusions about states that no meter reads, drawn from the meters, the topology and the models of the components. Table 3.2 lists these quantities by kind.

Table 3.2: Electrical quantities of a data center by the way they are obtained.

Quantity	Kind	Obtained from
Voltage V_i, current I_i, power P_i at a meter	measured	switchgear, UPS, PDU, tap-box and rack PDU meters
UPS and transformer losses	calculated	input minus output meters, or an efficiency curve and the measured throughput
Circuit loading	calculated or inferred	measured current over rating where metered; otherwise assembled from downstream meters through the topology
Actual load path	inferred	meter readings compared with the predictions of alternative topologies
Transfer-switch state	inferred	reported position where trusted; otherwise the pattern of path loads
Equipment availability	inferred	alarms, test records and recent behavior of the module or generator
Effective remaining capacity	inferred	ratings, inferred availability and inferred topology, evaluated after a credible failure

The transfer switch shows why the distinction matters. Suppose an illustrative 2 MW block of single-corded cooling equipment is fed through an automatic transfer switch that can select path A or path B. The operator's display shows it on A; physically it is on B, because a transfer during a test was never reversed or the position contact is faulty. Planned maintenance on path B is then assessed as affecting only the B cords of the dual-corded IT load, which the A cords cover. In fact taking B out of service also forces the cooling block through a transfer to A: an interruption of the block for the transfer time, an operation of a switch that may fail to operate, and an additional 2 MW on path A at the moment it is carrying the whole IT load. The operator's picture assigns this event no risk at all; the physical system assigns it a real one. The evidence for the true position is in the measurements, since path B's meter reads 2 MW more and path A's 2 MW less than the recorded position predicts, but it is evidence only to a model that treats the switch position as unknown rather than as given.

3.6.3 A reliability model of two paths

A dual-corded server is lost only if both of its paths fail. Write A_{f} and B_{f} for the events that path A and path B are failed at a given moment. The outage probability is

$$\Pr(\text{outage}) = \Pr(A_{\mathrm{f}} \cap B_{\mathrm{f}}) = \Pr(A_{\mathrm{f}})\,\Pr(B_{\mathrm{f}} \mid A_{\mathrm{f}}). \tag{3.13}$$

If the paths fail independently, the conditional probability equals the unconditional one, $\Pr(B_\mathrm{f} \mid A_\mathrm{f}) = \Pr(B_\mathrm{f})$, and with an illustrative $\Pr(A_\mathrm{f}) = \Pr(B_\mathrm{f}) = 0.01$,

$$\Pr(\text{outage}) = 0.01 \times 0.01 = 0.01^2 = 0.0001 = 0.01\,\%. \tag{3.14}$$

Duplication has reduced the outage probability by a factor of 100 compared with one path.

With a common cause, that calculation can be far too optimistic. A simple way to represent common causes is to suppose that a fraction β of each path's failure probability q comes from events that fail both paths together, and the remaining fraction $1 - \beta$ from independent events. To first order each path still fails with probability q, and, neglecting terms of order q^3,

$$\Pr(\text{outage}) \approx \beta q + \big((1 - \beta)\, q\big)^2. \tag{3.15}$$

With $q = 0.01$ and $\beta = 0.1$, one failure in ten being of the common-cause kind,

$$\Pr(\text{outage}) \approx 0.1 \times 0.01 + (0.9 \times 0.01)^2 = 0.001 + 0.000081 = 0.00108,$$

nearly eleven times the independent figure. The conditional probability of B failing given that A has failed is $0.00108/0.01 \approx 0.11$, not 0.01: knowing that one path has failed makes the failure of the other eleven times more likely. The first term dominates as soon as β exceeds about q, because a common cause needs one event while an independent double failure needs two.

Figure 3.7 plots equation (3.15). The horizontal axis is the common-cause fraction β; the vertical axis, on a logarithmic scale, is the outage probability of the two-path system. The solid curve is the formula with $q = 0.01$. The lower dashed line is the independent result of equation (3.14), and the upper dotted line is the failure probability q of a single path, the value with no redundancy at all. The curve leaves the independent value almost at once: at $\beta = 0.05$ it is already near six times higher, and at $\beta = 0.3$ it is thirty times higher, three quarters of the way, on the logarithmic scale, from the independent value back to the single-path value. Conditional dependence, not the count of paths, sets the reliability.

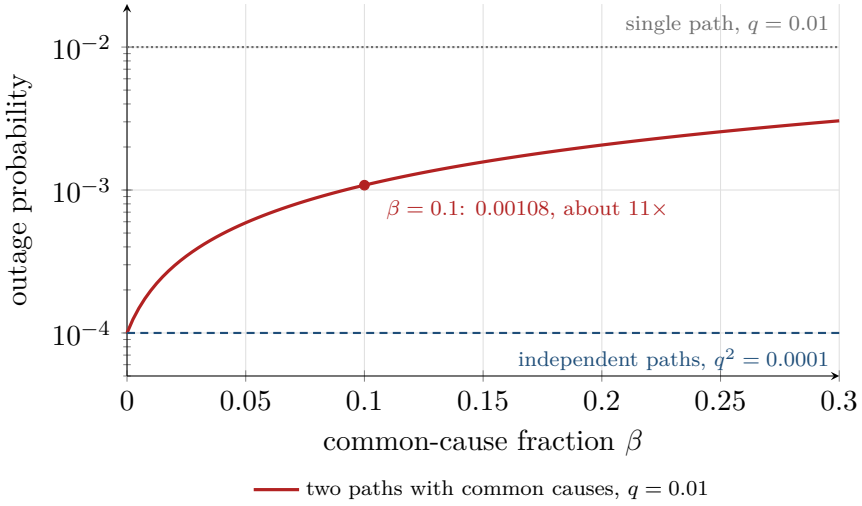


Figure 3.7: Outage probability of two redundant paths against the common-cause fraction, illustrative, $\Pr(\text{outage}) \approx \beta q + ((1 - \beta)q)^2$ (equation (3.15)) with path failure probability $q = 0.01$. At $\beta = 0$ the paths are independent and the probability is $q^2 = 0.0001$; any common cause raises it quickly toward the single-path value q.

Principle 3.7. *The reliability of redundant paths is set by their conditional dependence: a small fraction of common-cause failures raises the outage probability far above the product of the individual failure probabilities.*

3.6.4 The facility as a small power system

Taken together, the elements of this chapter form a small power system: buses (the switchgear, UPS and PDU busbars), transformers, feeders and busways, breakers, loads, backup generation, transfer schemes and a set of contingency states. It can be modeled the way a transmission system is. Collect the electrical state, the bus voltages and the power on every branch, in a vector x; the uncertain parameters, such as efficiencies, ratings and meter biases, in a vector θ; and the topology, the positions of breakers and transfer switches and the assignment of each rack to its PDU and path, in a discrete variable s. The internal network then satisfies a set of equations

$$F(x, \theta, s) = 0, \tag{3.16}$$

one power balance per bus, as in equation (3.10) for the main bus, and one relation per component, such as equation (3.9) for each UPS.

The meters give a vector y of measurements of some components of x, with errors. The questions of Sections 3.5.4 to 3.6.2 then become one question: which states, parameters and topologies are consistent with the measurements and with equation (3.16), and how probable is each? The answer is a joint probability distribution conditioned on the measurements,

$$p(x, \theta, s \mid y) \propto p(y \mid x, \theta, s)\, p(x, \theta, s), \qquad \text{subject to } F(x, \theta, s) = 0, \tag{3.17}$$

in which the measurement model $p(y \mid x, \theta, s)$ encodes the meters' accuracy and $p(x, \theta, s)$ encodes what is believed before the measurements, such as the recorded topology with a small probability of being wrong. The hypotheses H_1 to H_5 of Section 3.5.4 are regions of this space: a meter bias is a value of θ, a mislabeled rack a value of s. Estimating the state of such a network from its measurements is the subject of a later chapter.

3.6.5 A mental model in three layers

The equipment of this chapter falls into three layers. The source layer supplies energy: the utility grid, the generators and the batteries. The distribution layer carries and conditions it: switchgear, transformers, UPS, PDUs and busways. The load layer consumes it: racks, servers, the cooling plant and the auxiliaries. Four functions span all three layers rather than belonging to any one: protection, which isolates faults selectively; control, which starts generators, operates transfer switches and sets UPS modes; redundancy, which duplicates sources, paths and modules; and monitoring, which meters the flows and records the states.

Figure 3.8 draws the three layers as boxes from left to right in the direction of the power flow, with the components of each listed inside. The four horizontal bands below run under all three boxes, because each function acts on every layer: protection includes the utility's and generators' protection as well as the rack breakers; control reaches from the generator start signal to the server's power supply; redundancy is planned at every layer; and monitoring meters each of them. Placing a problem in this picture, by its layer and by the function involved, is the first step of diagnosing it.

3.6.6 From a chain to a network

The chain of Figure 3.1, utility to server, is the correct path of the power on any one day, but it is the wrong model of the system. The system is

$$\text{multiple sources} \rightarrow \text{multiple distribution paths} \rightarrow \text{critical loads},$$

with redundancy, protection and stored energy arranged so that the loads see continuous power through any single credible failure. The operational question it raises is accordingly not *whether the present load is below capacity*, which the margins of the present state answer, but *whether*

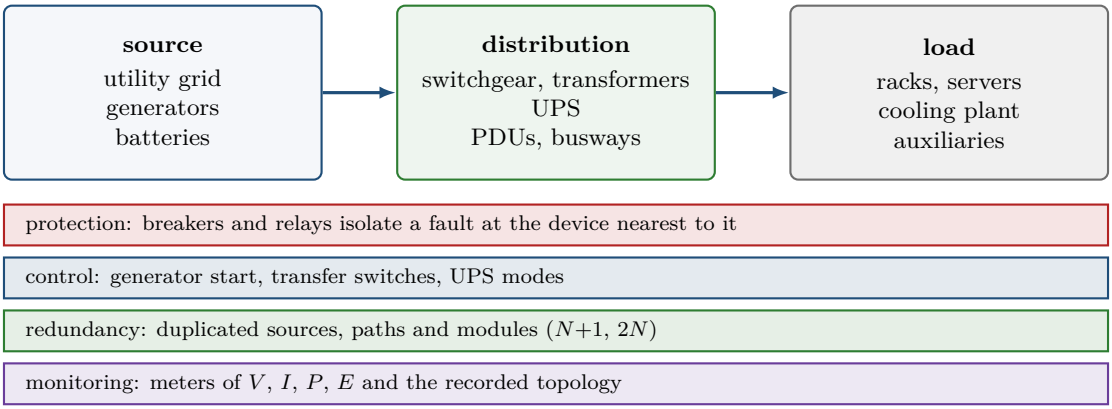


Figure 3.8: The power system of a data center in three layers, source, distribution and load, with the four functions that span all of them drawn as bands underneath.

the system can still support the load after a credible failure, which only the post-contingency state answers.

Because the post-contingency load depends on inferred quantities, the topology, the transfer-switch states, the availability of modules and the true values of metered loads, it is not known exactly, and the question is properly a probability,

$$\Pr(P_{\text{post-contingency}} > C_{\text{available}} \mid y), \tag{3.18}$$

the probability, given the measurements, that the load on a surviving path after a credible failure exceeds the capacity actually available. As an illustration, if the post-contingency load of a path rated 10 MW is estimated at 9.6 MW with a standard deviation of 0.4 MW from meter errors and topology uncertainty, and the estimate is approximately normal, the probability of overload is $\Pr(Z > (10 - 9.6)/0.4) = \Pr(Z > 1) \approx 0.16$ for a standard normal Z: a path that would pass a deterministic check, $9.6 < 10$, carries about a one-in-six chance of being overloaded. Reducing the standard deviation to 0.2 MW, by better metering or a verified topology, lowers it to $\Pr(Z > 2) \approx 0.023$. The treatment of capacity as a risk of this kind is developed in a later chapter.

Principle 3.8. *The operational question of a data center's power system is the probability that the load after a credible failure exceeds the capacity then available, not whether the present load is below the present capacity.*

3.7 What to retain

Definition 3.1 (Data center power distribution)**.** A data center's power system is a small, redundant power network, multiple sources feeding multiple protected paths to critical loads with stored energy bridging the start of backup generation, whose adequacy is judged by the probability that the load after a credible failure exceeds the capacity then available.

Table 3.3 lists the chapter's objects and the subsection that defines each.

3.8 Exercises

Exercise 3.1. A feeder carries 3 MW at a power factor of 0.9. (a) Using equation (3.3), compute the line current at 13.8 kV and at 480 V. (b) By what factor is the conductor loss larger at 480 V for the same conductors? (c) With a resistance of 0.005 Ω per conductor and three conductors, compute the loss at each voltage in kilowatts and as a percentage of the power carried.

Table 3.3: The objects of the power distribution model and where each is defined.

Object	Expression	Role	Defined in
Power path	utility to server, Figure 3.1	stages in the order power meets them	3.1.1
Selective protection	nearest device clears the fault	confines a fault to one circuit	3.2.2
Three-phase power	equation (3.3)	current for a given power and voltage	3.2.3
Conductor loss	$P_{\text{loss}} = I^2 R$	why power travels at high voltage	3.2.3
UPS energy	$E = P\,t$	minimum stored energy for the bridge	3.2.5
Double conversion	AC to DC to AC	isolates the load from the input	3.2.6
A and B paths	Figure 3.3	survive the loss of either path	3.3.4
N, $N{+}1$, $2N$	Figure 3.4	levels of duplication	3.4.1
Contingency capacity	$C_{\text{B}} \geq$ full load	rating after a path is lost	3.4.3
Capacity margin	$M = C - P$, $M' = C - P'$	present and post-contingency headroom	3.4.4
UPS loss	equation (3.9)	loss that becomes cooling demand	3.5.2
Facility power balance	equation (3.10), residual r	consistency test of the metering	3.5.3
Hypotheses H_1–H_5	Table 3.1	explanations of a residual	3.5.4
Two-path outage	equation (3.15)	reliability with common causes	3.6.3
Network model	$F(x, \theta, s) = 0$	the facility as a power system	3.6.4
Contingency risk	equation (3.18)	the operating question	3.6.6

Exercise 3.2. A UPS must carry a critical load of 3 MW for 8 min. (a) Compute the ideal stored energy. (b) With an inverter efficiency of 0.96, a usable fraction of 0.85 of nameplate energy and an end-of-life capacity of 0.8 of the new value, compute the nameplate energy required. (c) The UPS has an efficiency of 0.96 at this load. Compute its input power, its loss, and the energy lost over a year of continuous operation. Into which term of equation (3.10) does that loss go, and which other term does it raise?

Exercise 3.3. A critical load of 6 MW is supplied by UPS modules of 1.5 MW. (a) How many modules, and how many megawatts installed, do N, $N{+}1$ and $2N$ require, and at what fraction of its rating does each module run at full load? (b) In a $2N$ design, paths A and B are each rated 8 MW and share the 6 MW equally. Compute each path's margin in normal operation and after the loss of the other path. (c) The load grows to 8.5 MW, still shared equally. Is the normal state feasible? Is the post-contingency state? (d) For a third path rated 10 MW, the post-contingency load is estimated as approximately normal with mean 9.4 MW and standard deviation 0.3 MW. Compute the probability of overload after the contingency.

Exercise 3.4. (a) Two paths each fail with probability $q = 0.02$. Compute the outage probability if they are independent, and with a common-cause fraction $\beta = 0.05$ using equation (3.15). By what factor does the common cause raise it, and what is the probability that B is failed given that A is failed? (b) A facility's utility meter reads 15.0 MW and its measured and calculated loads sum to 14.4 MW. Compute the residual. The five hypotheses of Table 3.1 are given prior probabilities 0.30, 0.25, 0.15, 0.20 and 0.10, and a week of residuals, roughly constant day and night and uncorrelated with outdoor temperature or IT load, has likelihoods 0.2, 0.6, 0.1, 0.05 and 0.3 under H_1 to H_5. Compute the posterior probability of each hypothesis with equation (3.12), and name the measurement from Table 3.1 that would best test the leading one.

3.9 Connection to the TERRA framework

The power system of a data center is the first of the three conservation networks of the facility, and the one closest to a conventional power-system model.

- **Buses are conservation nodes, feeders are edges.** The switchgear, UPS and PDU busbars are nodes at which power balances, and the feeders, busways and cords are edges carrying power flows. Equation (3.10) is the balance of the main-bus node, Figure 3.6 is that node drawn by hand, and equation (3.16) is the set of all such balances.

- **Components carry efficiencies, ratings and states.** A UPS is a component with an efficiency relation, equation (3.9), a power rating and a stored-energy state; a transformer has its loss relation, a breaker or transfer switch a discrete position. The loss each component dissipates is the source term that couples the electrical network to the thermal network of the rooms it sits in.

- **The facility balance is a redundancy relation.** With the utility meter and the submeters all in the model, the balance has more measurements than unknowns, and its residual r is evidence. The hypotheses H_1 to H_5 become a ranked posterior, equation (3.12), and Table 3.1 is the list of next measurements ranked by what each would add.

- **Topology is an uncertain object.** Which rack is on which PDU and which path, and where each transfer switch sits, are the discrete variable s of equation (3.17), with the recorded wiring as a prior rather than as a fact. A mislabeled rack is then a posterior over assignments, detected by the paired residuals of Section 3.6.1.

- **Contingency risk is a probability.** Removing a path from the graph and simulating the post-contingency flows gives P' on each surviving edge; propagating the posterior uncertainty of measurements, parameters and topology through that simulation gives equation (3.18), the probability that the required power exceeds the available capacity.

In summary, a data center's power system is a redundant network of sources, protected paths and loads whose purpose is continuity; its losses close a facility balance that tests the metering, its reliability is set by the dependence between nominally independent paths, and its adequacy is the probability that a credible failure leaves a load on a path without enough capacity. The next chapter follows the critical transition of that network in time, from the loss of the utility through the UPS battery to the generators.

Chapter 4

UPS, batteries and backup power: the transition from grid to battery to generator

4.1 Two roles in one transition

Chapter 3 followed electrical power from the utility connection to the racks on the assumption that the utility is there. This chapter removes that assumption. A utility supply can sag for a fraction of a second during a fault elsewhere on the network or fail for hours during a storm, and a data center is built so that its IT equipment notices neither. The question of the chapter is

how the facility keeps its IT equipment powered through a grid disturbance or outage, and what determines whether it succeeds.

Two devices answer it, and they play different roles. The battery of the uninterruptible power supply (UPS) handles the immediate interruption: it supplies the load in the first instant, before anything mechanical can move. The standby generator handles the longer outage: it can run for hours or days on stored fuel, but it needs seconds to start and cannot respond at once. The chapter follows the transition between the two, sizes the battery that bridges it, describes the generator and the switches that connect it, ranks the loads that must be kept when capacity is short, and then treats the whole chain as what it is in operation: a system whose success is uncertain and is best stated as a probability.

Jargon: ride-through, runtime, double conversion, transfer switch, static transfer switch, VRLA, state of health, common-cause failure, load shedding

Ride-through is the ability of a load to keep operating through a disturbance of its supply; a UPS gives the IT load ride-through of a utility failure. *Runtime* (or autonomy) is the time a UPS battery can carry a given load from a given state of charge before it is depleted. *Double conversion* is the UPS topology in which the incoming alternating current (AC) is converted to direct current (DC) and back to AC, so that the load is always supplied by the output converter and the battery sits on the DC link between the two conversions. A *transfer switch* selects which of two sources, for example the utility or a generator, feeds a load; an *automatic transfer switch* (ATS) does so on its own when it detects that the preferred source has failed. A *static transfer switch* (STS) is a transfer switch built from semiconductor switches instead of moving contacts, so that it can move a load between two AC sources within a fraction of one AC cycle. *VRLA* stands for valve-regulated lead-acid: a sealed lead-acid battery whose internal gas is recombined and which vents only through a pressure valve. The *state of health* (SOH) of a battery is the ratio of its present usable capacity to its capacity when new. A *common-cause failure* is the failure of several components from one shared cause, such as a fuel supply that feeds all generators. *Load shedding* is the deliberate disconnection or reduction of lower-priority loads so that the available supply covers the higher-priority ones.

Key idea

The battery provides continuity; the generator provides duration. Whether the critical load stays powered is not settled by comparing a battery's capacity with a required energy: it depends jointly on the state of charge and health of the battery, the load, the converter efficiency, the generator start, the fuel and the switching, and it is properly stated as the probability that enough capacity remains available throughout the disturbance.

This section lays out the sequence of events that follows a utility failure and explains why the UPS can take the load without any delay.

4.1.1 The outage timeline

A utility outage unfolds in five steps. At $t = 0$ the utility fails. The UPS carries the load immediately, from its battery. The generator receives a start signal and runs through its start sequence. The generator stabilizes at rated speed, frequency and voltage. The load is transferred to the generator, and the battery stops discharging. Two of these steps happen at very different speeds. The UPS acts essentially instantaneously, because, as the next subsection shows, its battery is already on the power path. The generator cannot: it is a combustion engine coupled to an alternator, and it has to be cranked, accelerated and brought to the correct frequency and voltage before it can carry any load. The battery bridges the gap between the two.

Figure 4.1 draws the timeline for an illustrative load of 5 MW. The horizontal axis is the time in seconds relative to the utility failure; the vertical axis is the power each source delivers to the load, in megawatts. The gray curve is the utility, the green curve the UPS battery and the blue curve the generator. The shaded bands and the labels above the axis mark the four phases. Before $t = 0$ the utility supplies the whole load. At $t = 0$ its contribution falls to zero and the battery's rises to 5 MW in the same instant, so the load sees no gap. From 0 to 8 s the battery alone carries the load while the generator starts. From 8 to 18 s the load moves gradually from the battery to the generator: the UPS raises the power it draws from its generator-fed input over about ten seconds, a ramp often called *walk-in*, because a 5 MW step applied at once would pull the generator's speed, and with it its frequency and voltage, down. After 18 s the generator carries the load and the battery is idle. At every instant the three contributions add to 5 MW.

The battery's part in this nominal sequence is small. It delivers 5 MW for 8 s, then on average 2.5 MW for the 10 s ramp, which is

$$E = 5\,\text{MW} \times 8\,\text{s} + 2.5\,\text{MW} \times 10\,\text{s} = 40 + 25 = 65\,\text{MJ} = \frac{65}{3.6}\,\text{kWh} \approx 18\,\text{kWh}.$$

Section 4.2.1 sizes the same battery for ten minutes of full load, 833 kWh; the nominal transfer uses about 2 % of it. The rest of the energy is not there for the nominal sequence. It is there for the sequences that do not go as planned, which Section 4.3.2 lists.

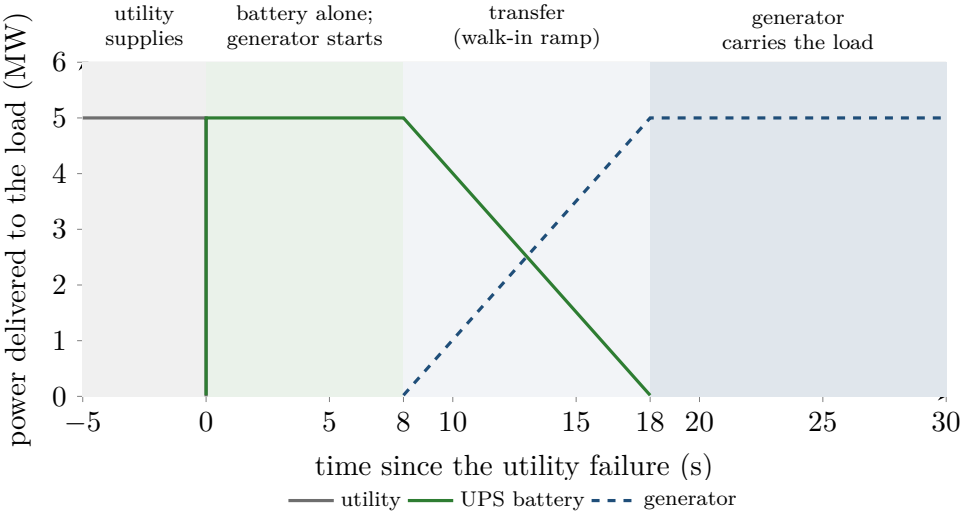


Figure 4.1: The outage timeline for an illustrative 5 MW load: power delivered by the utility, the UPS battery and the generator against time. The utility fails at $t = 0$, the battery takes the load in the same instant, the generator completes its start sequence at 8 s and takes the load over a 10 s ramp. The three contributions add to 5 MW at every instant; the battery delivers about 18 kWh in all.

4.1.2 Why a double-conversion UPS responds at once

The speed of the UPS is a consequence of its topology. In a double-conversion UPS the incoming AC from the utility passes through a *rectifier*, which converts it to DC, onto a *DC bus* (the DC link), and from the DC bus through an *inverter*, which converts it back to AC for the load:

$$\text{AC from the utility} \rightarrow \text{rectifier} \rightarrow \text{DC bus} \rightarrow \text{inverter} \rightarrow \text{AC to the load.}$$

The battery is connected to the DC bus. The load is supplied through the inverter at all times, in normal operation as well as during an outage, and the inverter does not know or care where the DC on its input comes from.

Figure 4.2 draws the path in its two modes. Each panel reads from left to right: the utility and the generator on the left feed an automatic transfer switch, which feeds the rectifier; the rectifier feeds the DC bus; the battery hangs below the DC bus; the inverter takes the DC bus to the IT load on the right. Thick arrows are the paths that carry power in that mode; dashed gray arrows are paths that are connected but idle. In panel (a), normal operation, power flows from the utility through the rectifier and the inverter to the load, and a small flow from the DC bus keeps the battery fully charged (a *float charge*); the generator stands by. In panel (b), the first seconds of an outage, the rectifier's output has fallen to zero. The battery was already connected to the DC bus and already at the bus voltage, so the moment the rectifier stops supplying current, the battery does: current reverses in the battery's connection without any

switch opening or closing, and the inverter continues to see DC on its input. No mechanical device has to start and no contact has to move. In symbols, normal operation is

$$P_{\text{grid}} \;\rightarrow\; P_{\text{load}}, \qquad \text{battery held at full charge,} \tag{4.1}$$

and an outage, with $P_{\text{grid}} = 0$, is

$$P_{\text{bat}} \;\rightarrow\; P_{\text{inv}} \;\rightarrow\; P_{\text{IT}}. \tag{4.2}$$

Two other features of the topology matter later. First, because the load always passes through both conversions, a double-conversion UPS has conversion losses all the time, not only during an outage; Section 4.2.4 quantifies the inverter's share and where its heat goes. Second, the inverter output is regenerated from DC, so it is isolated from disturbances on the utility side: a voltage dip that lasts a few cycles reaches the DC bus as a brief contribution from the battery and does not reach the load at all. Simpler UPS designs that keep the load on the utility in normal operation and switch to the inverter only when the utility fails exist, but they need a transfer time; the double-conversion design is the usual choice for critical IT load because it does not.

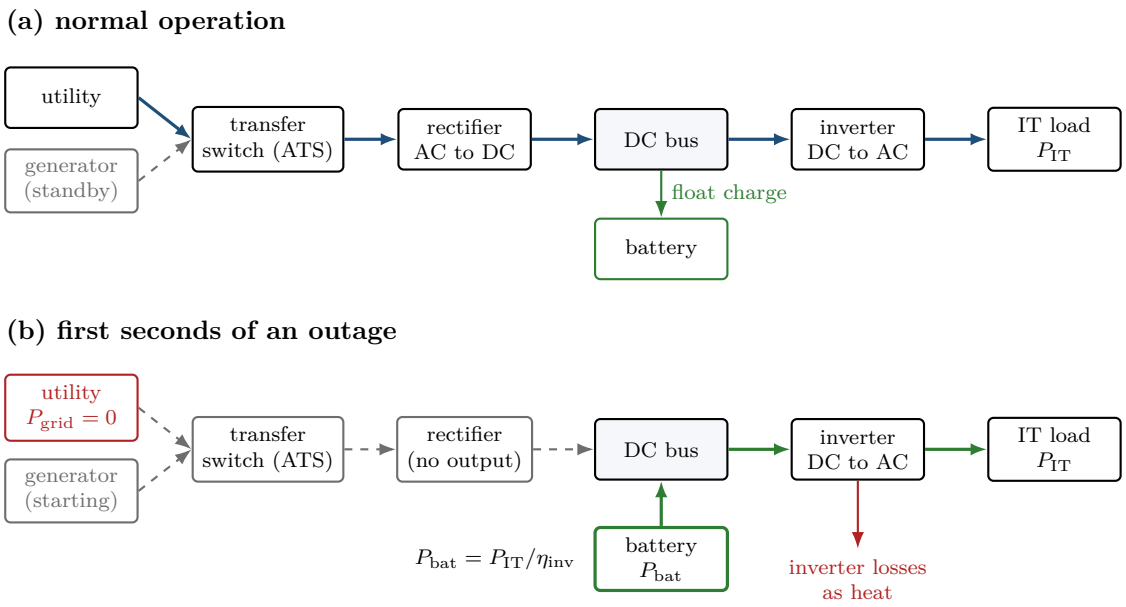


Figure 4.2: The double-conversion UPS path. (a) Normal operation: power flows from the utility through the transfer switch, the rectifier, the DC bus and the inverter to the IT load; the battery is held at full charge from the DC bus. (b) Outage: the rectifier has no input, and the battery, already on the DC bus, supplies the inverter without any switching. Thick arrows carry power; dashed gray arrows are connected but idle.

Principle 4.1. *A double-conversion UPS responds at once because its battery is always on the power path: when the rectifier stops supplying the DC bus, the battery does, with nothing to start and nothing to switch.*

4.2 Sizing the battery

The battery has to supply the load for long enough, at a high enough power, with the losses of the path in between. This section computes the energy requirement for an example, separates the power requirement from the energy requirement, writes the state of charge as the state of a balance, and adds the inverter efficiency and its heat.

4.2.1 The energy requirement

Suppose the critical load is $P_{\text{load}} = 5\,\text{MW}$ and the battery must carry it for 10 minutes. The energy is power times time,

$$E = P_{\text{load}}\, t = 5\,\text{MW} \times \frac{10}{60}\,\text{h} = 0.833\,\text{MWh} = 833\,\text{kWh}. \tag{4.3}$$

This is the ideal requirement: the energy the load itself consumes. The battery that is installed must be larger, for six reasons. The *inverter losses* of Section 4.2.4 mean the battery delivers more than the load receives. The *battery's own efficiency* is below one: part of the stored chemical energy is lost as heat inside the cells, more so at the high discharge rates of a ten-minute runtime. *Aging* reduces the usable capacity over the battery's life, and the battery must still meet the requirement at the end of that life, not only when new. The *allowable depth of discharge*, the fraction of the stored energy that may be withdrawn before a voltage limit is reached or the cells are damaged, is below one. A *reserve margin* covers load growth and estimation error. And *temperature effects* reduce the deliverable energy of a cold battery.

Table 4.1 applies one illustrative factor for each reason in turn. The left column names the reason, the middle column the factor applied, and the right column the energy after applying it. Each loss divides the requirement by an efficiency or a fraction below one; the margin multiplies it.

Table 4.1: From the ideal energy of equation (4.3) to an installed nominal capacity, with illustrative factors. Each row applies one factor to the row above.

Reason	Factor (illustrative)	Energy (kWh)
Ideal load energy, 5 MW for 10 min	–	833
Inverter losses	÷ 0.96	868
Battery internal losses at high discharge rate	÷ 0.97	895
Aging: requirement met at end of life, SOH = 0.8	÷ 0.80	1119
Allowable depth of discharge	÷ 0.90	1243
Reserve margin	× 1.10	1367
Temperature derating	÷ 0.95	1439

The illustrative installed capacity is about 1.44 MWh, 1.73 times the ideal 833 kWh. No single factor is large; their product is. The largest single step is aging, and it is the one that changes during operation: Section 4.5.2 shows what happens when the state of health actually reached falls below the value assumed in the design.

4.2.2 Power versus energy

A battery must satisfy two separate requirements: enough *power* to support the load at every instant, and enough *energy* to sustain it for the required time. They are linked by

$$E = P\, t, \tag{4.4}$$

but neither one implies the other. A battery with $E = 1\,\text{MWh}$ says nothing about whether it can deliver 10 MW: that depends on its converters, its cell chemistry and its internal resistance. If it could, it would do so for only $1/10\,\text{h} = 6$ minutes. Conversely, a rating of $P^{\max} = 10\,\text{MW}$ says nothing about duration: 10 MW for 10 seconds needs $10\,000\,\text{kW} \times 10/3600\,\text{h} = 27.8\,\text{kWh}$, and 10 MW for 30 minutes needs 5 MWh, a factor of 180 more.

A UPS battery sits at an unusual point on this spectrum. The example of Section 4.2.1 asks for 5 MW from 833 kWh (ideally), a power-to-energy ratio of $5/0.833 = 6$ per hour, so the battery is emptied in a sixth of an hour. Battery systems built to shift energy on a power grid typically run at ratios of 0.25 to 1 per hour, emptying in one to four hours. A UPS battery is a high-power, short-duration device, and its cells, its cooling and its protection are chosen for that: high discharge currents for a few minutes, not moderate currents for hours.

4.2.3 State of charge

During a discharge the stored energy E falls at the rate the battery delivers power,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -P_{\mathrm{bat}}, \tag{4.5}$$

a one-line conservation law for the energy in the battery. The state of charge is the stored energy as a fraction of the usable energy,

$$\mathrm{SOC} = \frac{E}{E_{\mathrm{usable}}}, \qquad \frac{\mathrm{dSOC}}{\mathrm{d}t} = -\frac{P_{\mathrm{bat}}}{E_{\mathrm{usable}}}, \tag{4.6}$$

where the battery's internal efficiency is ignored, as it is in the rest of this chapter except where stated. For a constant battery power the state of charge falls linearly,

$$\mathrm{SOC}(t) = \mathrm{SOC}_0 - \frac{P_{\mathrm{bat}}\, t}{E_{\mathrm{usable}}}, \qquad t_{\mathrm{depl}} = \frac{\mathrm{SOC}_0\, E_{\mathrm{usable}}}{P_{\mathrm{bat}}}, \tag{4.7}$$

and reaches zero at the depletion time t_{depl}. The state of charge is a state in the sense of a dynamic model: its present value summarizes the whole history of charge and discharge, and its rate of change is set by the present power.

Figure 4.3 plots equation (4.7) for an illustrative battery with $E_{\mathrm{usable}} = 1$ MWh, starting full, discharging at $P_{\mathrm{bat}} = 4.17$ MW (a 4 MW load through an inverter of efficiency 0.96, Section 4.2.4). The horizontal axis is minutes since the utility failure; the vertical axis is the state of charge. The gray line is the discharge if no generator ever takes the load: the battery empties at $t_{\mathrm{depl}} = 1000/4167\,\mathrm{h} = 0.24\,\mathrm{h} = 14.4$ minutes. The solid green line is the nominal case of Figure 4.1: the generator takes the load at about 8 s, the discharge stops, and the state of charge stays at $1 - (8/60)/14.4 = 0.991$. The dashed blue line is a case in which the generator's first start attempt fails and a second attempt succeeds at 2 minutes (illustrative): the discharge stops at $1 - 2/14.4 = 0.861$. The brace marks the window between the nominal generator start and depletion. Every minute of runtime beyond the nominal start is time in which a failed start can be recovered; the shape of the curve says that each minute costs the same $1/14.4 = 7\,\%$ of charge.

4.2.4 Inverter efficiency and the heat it produces

The inverter between the DC bus and the load is not lossless. With an inverter efficiency η_{inv}, the battery must supply

$$P_{\mathrm{bat}} = \frac{P_{\mathrm{load}}}{\eta_{\mathrm{inv}}}, \tag{4.8}$$

so that for $\eta_{\mathrm{inv}} = 0.96$ and $P_{\mathrm{load}} = 4$ MW,

$$P_{\mathrm{bat}} = \frac{4\,\mathrm{MW}}{0.96} = 4.17\,\mathrm{MW}.$$

The difference, 0.17 MW or about 170 kW, does not disappear. It becomes heat in the inverter's semiconductors, magnetics and capacitors, and that heat is released into the UPS room. The

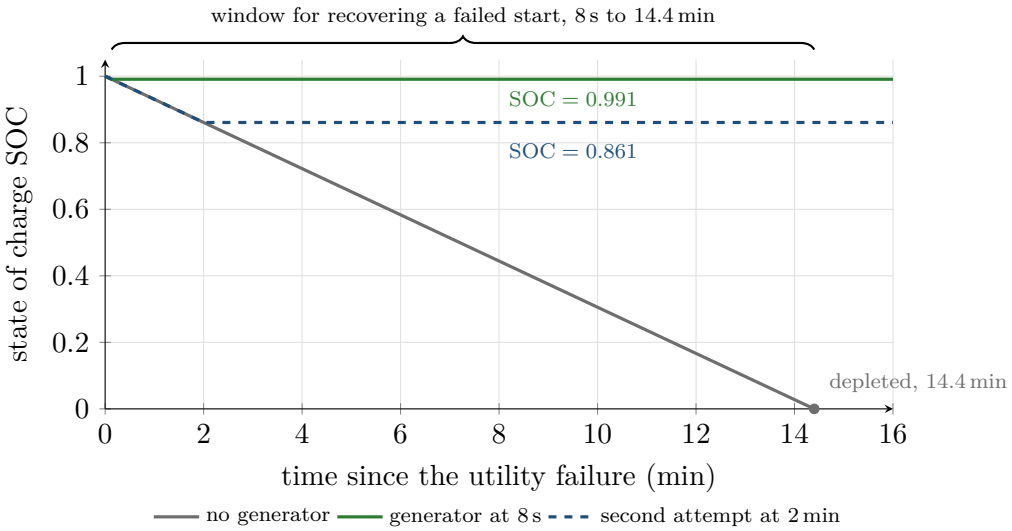


Figure 4.3: State of charge during a discharge, $\mathrm{SOC}(t) = 1 - P_{\mathrm{bat}} t / E_{\mathrm{usable}}$ (equation (4.7)), with illustrative $E_{\mathrm{usable}} = 1\,\mathrm{MWh}$ and $P_{\mathrm{bat}} = 4.17\,\mathrm{MW}$, so the battery empties in 14.4 minutes. The discharge stops when the generator takes the load: at 8 s in the nominal case, at 2 minutes after a failed first attempt. Recharging after the transfer is not drawn.

UPS room therefore needs its own cooling, sized for the converter losses at full load, and that cooling has to keep running during an outage, because the losses are largest exactly when the UPS is carrying the load from its battery at full power. In normal operation the rectifier adds its own loss on top, so a double-conversion UPS is a continuous heat source of a few percent of the IT load.

The general point recurs throughout the book: an electrical inefficiency is a thermal load. Every watt lost in conversion appears as a watt of heat that some cooling system must remove, which in turn consumes electrical power. In the energy sizing of Section 4.2.1 the same factor appears as the second row of Table 4.1: the 4 % inverter loss raises the energy the battery must hold from 833 to 868 kWh.

Principle 4.2. *A backup battery must meet a power requirement and an energy requirement separately; its state of charge falls at the rate $P_{\mathrm{bat}}/E_{\mathrm{usable}}$, with P_{bat} larger than the load by the inverter loss, and that loss is heat to be removed.*

4.3 The generator and the transfer

The generator supplies the outage after the first seconds, and it is connected by switches that decide which source feeds which load. This section describes the start sequence, explains why the battery is sized for minutes rather than seconds, counts the generators a load requires, estimates how long the fuel lasts, and describes the automatic and static transfer switches.

4.3.1 The generator start sequence

A standby generator is a diesel or gas engine driving a synchronous alternator. From a start signal it must:

1. *crank*: a starter motor, powered by a dedicated starting battery or by compressed air, turns the engine until it fires;
2. *reach speed*: the engine accelerates to its rated speed, for example 1800 revolutions per minute for a four-pole alternator on a 60 Hz system;

3. *establish frequency*: the speed governor holds the speed, and with it the output frequency, at the rated value;
4. *establish voltage*: the excitation system builds the alternator's magnetic field until the output voltage is at the rated value;
5. *synchronize, if necessary*: when several generators share one bus, each must match the voltage, frequency and phase angle of the bus before its breaker can close;
6. *connect to the bus*: the generator breaker closes and the generator is available to take load.

If this sequence takes 8 s, the battery must hold the load for at least 8 s. That is the hard lower bound on runtime. A design figure of about ten seconds is common for standby generators in critical facilities; the example of this chapter uses 8 s. Designers provide far more battery than the start time requires: the battery of Section 4.2.1 holds ten minutes, seventy-five times the eight seconds. The next subsection explains why.

4.3.2 Runtime as a decision window

The start sequence can fail, and the extra runtime exists for the cases in which it does. The failures that occur in practice include:

- the generator fails to crank (a weak starting battery, a failed starter motor);
- a fuel problem (contaminated fuel, a clogged filter, air in the fuel line, a closed valve);
- the generator breaker fails to close;
- the transfer switch fails to transfer;
- the generator starts but its voltage or frequency is unstable and the protection will not let it connect;
- in a multi-generator plant, one generator starts and another does not, so the plant has less capacity than the load;
- a generator is out of service for maintenance when the outage arrives.

Each of these converts an eight-second event into one that needs a response. The runtime is the window in which that response can happen: the controls can retry the start, operators can transfer load to another generator or bus, noncritical loads can be shed so that the generators that did start can carry the rest, and, if none of that succeeds, the IT systems can be shut down in a controlled way, saving their state, instead of losing power abruptly.

Runtime is therefore a *risk buffer*, not only a margin over the nominal transfer time. A typical choice of 5 to 15 minutes reflects how long the recovery actions above take, not how long the generator takes to start. Figure 4.3 shows the trade: each minute added to the window costs battery energy, and the energy is paid for in every year of the battery's life, whether or not a start ever fails.

> **Principle 4.3.** *Battery runtime is sized for the start sequences that fail, not for the one that succeeds: the minutes beyond the nominal transfer time are the window in which a failed start is retried, load is shed or the IT load is shut down in order.*

4.3.3 Generator capacity and the N+1 rule

The generator plant must carry the whole critical load, which includes the IT load and the cooling that keeps it alive. With a critical load $P_{\text{crit}} = 20\,\text{MW}$ and generators of $P_{\text{unit}} = 4\,\text{MW}$ each, the number of units required is

$$N = \left\lceil \frac{P_{\text{crit}}}{P_{\text{unit}}} \right\rceil = \left\lceil \frac{20}{4} \right\rceil = 5, \tag{4.9}$$

where $\lceil \cdot \rceil$ rounds up to the next integer. An N+1 design installs one more than required, here 6 units with 24 MW in all, so that the plant still covers the load when any one unit is unavailable,

whether it failed to start or is out for maintenance. N+1 protects against exactly one unit missing. It does not protect against two, and it does not protect against a cause that removes several units at once; Section 4.6.2 returns to that difference.

4.3.4 Fuel inventory and long-duration runtime

Once the generators carry the load, the battery is no longer the limit; the fuel is. With a stored fuel volume F_{stored} and a consumption rate $\dot{F}$, the generator runtime is

$$t_{\text{fuel}} = \frac{F_{\text{stored}}}{\dot{F}}. \tag{4.10}$$

For an illustrative plant carrying 20 MW with a specific consumption of about 0.26 liters of diesel per kilowatt-hour of output (a typical order for large diesel generators near full load), $\dot{F} = 20\,000\,\text{kW} \times 0.26 = 5200$ liters per hour, and a stored volume of 200 000 liters gives

$$t_{\text{fuel}} = \frac{200\,000}{5200} \approx 38 \text{ hours}.$$

The consumption rate is not a constant of the plant. It depends on the load level, because the plant burns more fuel at higher load, and on the generator efficiency at that load, because a diesel engine's fuel per kilowatt-hour rises at part load. Long-duration resilience therefore depends on four things together: the fuel inventory, the generator efficiency, the load level and the logistics of refueling, which during a regional event that caused the outage may be the hardest of the four.

The fuel system also shows a structural point. A facility can have electrically redundant generators, six units for a load that needs five, and still have all of them drawing from one fuel tank, one fuel pump set or one supply line. Electrical redundancy then sits on top of a single fuel dependency. Redundancy does not imply independence: two components are redundant if either can do the job, and independent only if they do not share a cause of failure.

4.3.5 Automatic transfer switches

An automatic transfer switch selects which source feeds a load, the utility or the generator, and changes the selection by itself. It monitors the voltage, the frequency and the phase of each source and judges whether each is available. The logic for a utility failure is a chain of conditions,

$$V_{\text{grid}} < V_{\text{min}} \rightarrow \text{utility declared unavailable} \rightarrow \text{generator start signal} \rightarrow \text{generator voltage and frequency within tolerance} \rightarrow \text{transfer to the generator}.$$

The threshold is usually applied with a short time delay, so that a dip of a few cycles, which the UPS rides through from its battery, does not start the generators. When the utility returns, the switch waits for it to be stable for a set time and then transfers back, after which the generators run unloaded for a cool-down period and stop. Every step in this chain is a decision made from measurements, and each can be wrong: a miscalibrated voltage sensor can start the generators for nothing or fail to start them when needed.

In the architecture of Figure 4.2 the ATS sits upstream of the UPS, so its transfer is invisible to the IT load: during the transfer the UPS input has no power, and the battery covers the gap.

4.3.6 Static transfer switches

A static transfer switch switches between two AC sources electronically, with semiconductor switches instead of moving contacts. Its typical use is between two UPS systems,

$$\text{UPS A} \rightarrow \text{STS} \leftarrow \text{UPS B} \qquad \Longrightarrow \qquad \text{load}\ ,$$

so that a load fed through it can be moved from one UPS to the other in a few milliseconds, a fraction of one AC cycle, short enough that a server power supply does not notice. The purpose is to let equipment with a single power input still benefit from A/B redundancy. Most servers have two power supplies and can take one from each of two independent paths; a device with a single input cannot, and an STS gives it the same protection at the cost of one more component that must itself work. A static switch moves load between sources that are already energized; it does not create energy and does not replace the battery or the generator.

4.4 Load priority and the thermal buffer

When the available supply is less than the total load, the facility does not have to fail; it has to choose what to keep. This section ranks the loads, works through one shedding decision, and shows why the cooling cannot be cut as freely as the electrical arithmetic suggests.

4.4.1 Load tiers

Not all load in a data center is equally important. A common ranking has four tiers:

1. *Tier 1, critical IT*: the computing, storage and network equipment the facility exists to keep running.
2. *Tier 2, cooling for critical IT*: the fans, pumps, chillers and controls without which tier 1 overheats within minutes.
3. *Tier 3, less critical computing*: workloads that can be paused, migrated to another site or run at reduced speed.
4. *Tier 4, nonessential building services*: office lighting, general ventilation, kitchens and similar loads.

The consequence of the ranking is that $P_{\text{available}} < P_{\text{total}}$ does not mean collapse. With load shedding the facility disconnects or reduces the lowest tiers first and preserves the most important functions with the capacity that remains. Tier 2 sits above tier 3 because cooling is not optional for the IT it serves: shed the cooling of tier 1 and tier 1 is lost anyway, a few minutes later.

4.4.2 Shedding 4 MW of 25

Suppose the facility requires $P_{\text{required}} = 25\,\text{MW}$ and the generators that started provide $P_{\text{gen}} = 21\,\text{MW}$. Then

$$P_{\text{required}} - P_{\text{gen}} = 25 - 21 = 4\,\text{MW}$$

must be shed or reduced. Three means are available: shed noncritical IT, reduce cooling auxiliaries, or cap the power of servers so that they run slower and draw less. Each has a cost. Shedding computing interrupts the services it runs. Power capping slows every workload a little instead of stopping some.

Figure 4.4 draws one illustrative split of the 25 MW: tier 1 at 15 MW, tier 2 at 5 MW, tier 3 at 3 MW and tier 4 at 2 MW. The vertical axis is power in megawatts. The left column stacks the four tiers as required; the right column shows what is supplied after shedding. The dashed red line is the available generation, 21 MW. Tiers 1 and 2 together need 20 MW, the critical load of Section 4.3.3, and fit under the line with 1 MW to spare. The shedding order follows the tiers: all of tier 4, 2 MW, then 2 MW of tier 3, which keeps 1 MW of tier 3 running. The outlined blocks above the line in the right column are the shed load.

The arithmetic of the figure treats the tiers as independent blocks. They are not. Reducing cooling auxiliaries is the one option that appears free in the short term, because the IT load keeps running, and it is also the one that can turn an electrical shortfall into a thermal one: if

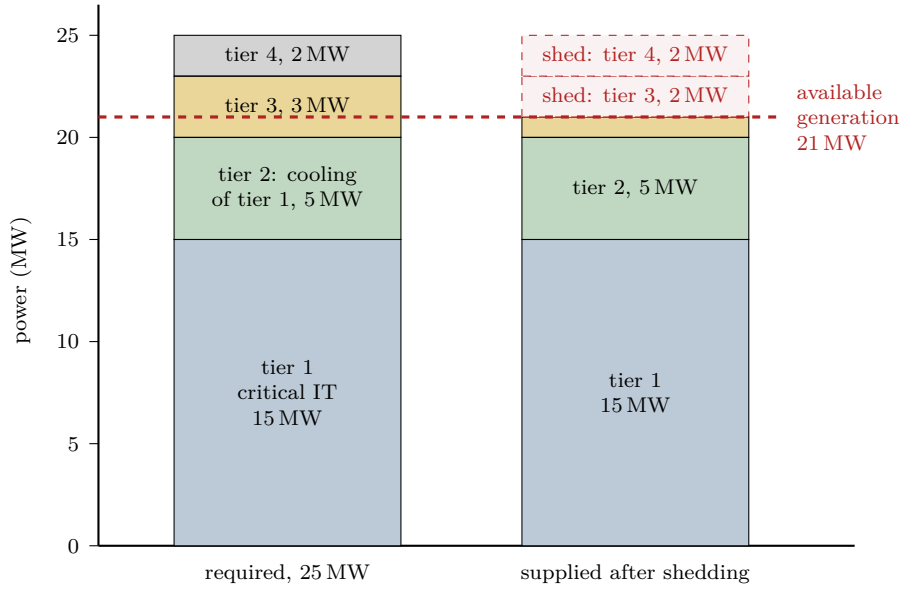


Figure 4.4: Load tiers against available generation, illustrative. The required 25 MW (left) is split into tier 1 (critical IT), tier 2 (cooling of tier 1), tier 3 (less critical computing) and tier 4 (nonessential building services). With 21 MW of generation (dashed), tier 4 and 2 MW of tier 3 are shed (right), and 1 MW of tier 3 remains in service.

the cooling of tier 1 is reduced too far, tier 1 overheats and is lost through its temperature limits instead of its power supply. The electrical and thermal systems are coupled, and a load-shedding decision that is feasible on the electrical side must also be feasible on the thermal side. The next subsection gives the time scale on which the thermal side responds.

4.4.3 The thermal buffer

During a transfer, cooling power can dip even when the IT load does not. Chillers may trip on the loss of utility power and take minutes to restart; fans and pumps on generator-fed boards lose power for the seconds of the transfer. The temperature of the room does not jump when that happens. Lumping the air, the equipment and the water that exchange heat with the IT load into a single thermal capacitance C_{th}, the energy balance is

$$C_{\mathrm{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\mathrm{IT}} - \dot{Q}_{\mathrm{cool}}, \tag{4.11}$$

where $\dot{Q}_{\mathrm{IT}}$ is the heat generated by the IT equipment, equal to its electrical power, and $\dot{Q}_{\mathrm{cool}}$ is the heat the cooling system removes. When $\dot{Q}_{\mathrm{IT}} > \dot{Q}_{\mathrm{cool}}$ the right-hand side is positive and $\mathrm{d}T/\mathrm{d}t > 0$: the temperature rises, but gradually, at a rate set by the size of the imbalance divided by the thermal capacitance.

How gradually depends on which thermal mass takes part. Take an illustrative data hall with $\dot{Q}_{\mathrm{IT}} = 1\,\mathrm{MW}$ and cooling that stops entirely, $\dot{Q}_{\mathrm{cool}} = 0$, for 30 s.

- If only the room air absorbs the heat, $5000\,\mathrm{m}^3$ of air at $1.2\,\mathrm{kg/m}^3$ and $c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ/(kg\,K)}$ give $C_{\mathrm{th}} = 6000 \times 1.0 = 6\,\mathrm{MJ/K}$, and $\mathrm{d}T/\mathrm{d}t = 10^6/(6 \times 10^6) = 0.17\,\mathrm{K}$ per second: 5 K in 30 s.
- If the IT equipment's own metal and the building surfaces in contact with the air take part, with an illustrative effective $C_{\mathrm{th}} = 30\,\mathrm{MJ/K}$, the rate is 0.033 K per second: 1 K in 30 s.
- If the chilled-water pumps and the air-handler fans stay on the UPS, so that the room air keeps passing over coils filled with circulating water, the water volume also takes part.

An illustrative $100\,\mathrm{m}^3$ of water at $c_{p,w} = 4.18\,\mathrm{kJ/(kg\,K)}$ adds $100\,000 \times 4.18 = 418\,\mathrm{MJ/K}$, and the rate falls to $10^6/(4.48 \times 10^8) = 0.0022\,\mathrm{K}$ per second, about $0.13\,\mathrm{K}$ per minute.

The comparison shows why tier 2 includes the fans and pumps and why they are often fed from a UPS: the thermal buffer of the water exists only while the water moves. A single lumped capacitance overstates how evenly the heat spreads, because hot spots at rack inlets respond faster than the average; the orders of magnitude are what the example is for. Chapter 8 develops the dynamics of these thermal masses.

Two kinds of stored energy. An outage is survived on stored energy of two kinds. *Electrical* storage in the UPS batteries keeps the IT equipment and the most important cooling auxiliaries powered. *Thermal* storage in the equipment, the air and the water, and in some facilities in tanks of chilled water built for the purpose, keeps temperatures within limits while the cooling plant is short of power or restarting. Both are buffers that buy time, both are depleted at a rate set by an imbalance, and both have to be counted when deciding how long a facility can ride through a disturbance.

Principle 4.4. *Electrical and thermal storage are both buffers that buy time during an outage. Shedding cooling to save electrical capacity draws down the thermal buffer, at a rate set by the heat imbalance divided by the thermal mass that the remaining fans and pumps keep connected.*

4.5 Battery health and runtime under uncertainty

The sizing of Section 4.2 assumed the battery's usable energy is known. It is not, and it changes with age. This section compares battery technologies, defines the state of health, turns runtime into a probability distribution, states the reliability question as a probability, and shows how a full state of charge can hide an insufficient battery.

4.5.1 Battery technologies

UPS batteries have historically been valve-regulated lead-acid (VRLA); lithium-ion batteries are increasingly used. Table 4.2 compares the two on the properties that matter for backup service. Each row is one property; the two right-hand columns give the typical behavior of each technology in general terms. The values are broad tendencies, not specifications of any product.

The last row matters most for reliability modeling. The two chemistries age by different mechanisms, at different rates, with different sensitivities to temperature and to cycling, and with different signatures in the measurements. A model that predicts the usable energy of a VRLA string from its age and room temperature is not a model of a lithium-ion system, and the monitoring row sets how much evidence about that aging the facility actually has.

4.5.2 Degradation and state of health

A battery's usable energy falls as it ages. With the state of health defined as the ratio of present to new capacity,

$$E_{\mathrm{usable}} = \mathrm{SOH} \cdot E_{\mathrm{nominal}}. \tag{4.12}$$

For $E_{\mathrm{nominal}} = 1\,\mathrm{MWh}$ and $\mathrm{SOH} = 0.8$ the battery holds about $E_{\mathrm{usable}} = 0.8\,\mathrm{MWh}$. A runtime computed from the nameplate is then no longer available. For the $5\,\mathrm{MW}$ load of Section 4.2.1 through an inverter of efficiency 0.96, the battery delivers $5/0.96 = 5.21\,\mathrm{MW}$, and

$$t = \frac{1000\,\mathrm{kWh}}{5208\,\mathrm{kW}} = 0.192\,\mathrm{h} = 11.5\,\mathrm{min} \quad \text{when new,} \qquad t = \frac{800}{5208}\,\mathrm{h} = 9.2\,\mathrm{min} \quad \text{at SOH} = 0.8.$$

Table 4.2: VRLA and lithium-ion batteries for UPS service: typical tendencies.

Property	VRLA (lead-acid)	Lithium-ion
Energy density	low; large floor area and weight per kWh	roughly two to three times higher by weight; smaller footprint
Lifetime	a few years in typical UPS service; shortened markedly by heat	often a decade or more
Thermal behavior	life falls steeply with room temperature above about 25°C	tolerates warmer rooms; risk of thermal runaway in fault conditions
Cost	lower purchase cost	higher purchase cost; fewer replacements over the facility life
Maintenance	periodic inspection and testing; whole strings replaced	less routine maintenance; depends on the battery management system
Safety	hydrogen release on overcharge; low runaway risk	thermal runaway must be prevented by monitoring, protection and spacing
Monitoring	often string voltage and current only, sometimes block voltages	battery management system reports cell voltages, temperatures and estimates of state of charge and health
Degradation	grid corrosion, sulfation, dry-out of the electrolyte	loss of lithium inventory, loss of active material, rising resistance

A battery that met a ten-minute requirement with a margin when new no longer meets it at the state of health that is often used to define end of life. This is why the sizing of Table 4.1 divides by the end-of-life state of health: a battery sized for its first day fails its requirement long before it is replaced.

4.5.3 Runtime as a distribution

With a known usable energy the runtime is one number. For $E_{\text{usable}} = 600\,\text{kWh}$ and a battery power $P = 3\,\text{MW}$,

$$t = \frac{E_{\text{usable}}}{P} = \frac{600\,\text{kWh}}{3000\,\text{kW}} = 0.2\,\text{h} = 12\,\text{min}. \tag{4.13}$$

The usable energy is not known exactly. It is not measured: it is inferred from the battery's age, its temperature history, its discharge tests and its voltage behavior, each with uncertainty. Suppose the best statement available is $E_{\text{usable}} = 600 \pm 100\,\text{kWh}$, read as a normal distribution with mean 600 kWh and standard deviation 100 kWh. Because runtime is the usable energy divided by a fixed power, it inherits the distribution with every value divided by 3000 kW:

$$E_{\text{usable}} \sim \mathcal{N}(600,\, 100^2)\ \text{kWh} \quad\Longrightarrow\quad t \sim \mathcal{N}(12,\, 2^2)\ \text{min}. \tag{4.14}$$

The runtime is then better described by its distribution $p(t)$ than by a single value. A single value of 12 minutes hides the fact that the runtime is below 10 minutes with probability

$$\Pr(t < 10\,\text{min}) = \Phi\left(\frac{10 - 12}{2}\right) = \Phi(-1) = 0.159,$$

where Φ is the standard normal cumulative distribution function, and below 8 minutes with probability $\Phi(-2) = 0.023$.

Figure 4.5 plots the distribution. The horizontal axis is runtime in minutes, with the corresponding usable energy in kilowatt hours printed under each tick; the vertical axis is

probability density per minute. The bell curve is $p(t)$; the dashed vertical line is its mean, 12 minutes; the shaded area to the left of 10 minutes is the probability 0.159 that the runtime falls short of a ten-minute requirement. A design that states "12 minutes" and a requirement of "10 minutes" looks safe by two minutes. The distribution says the requirement is missed about one time in six.

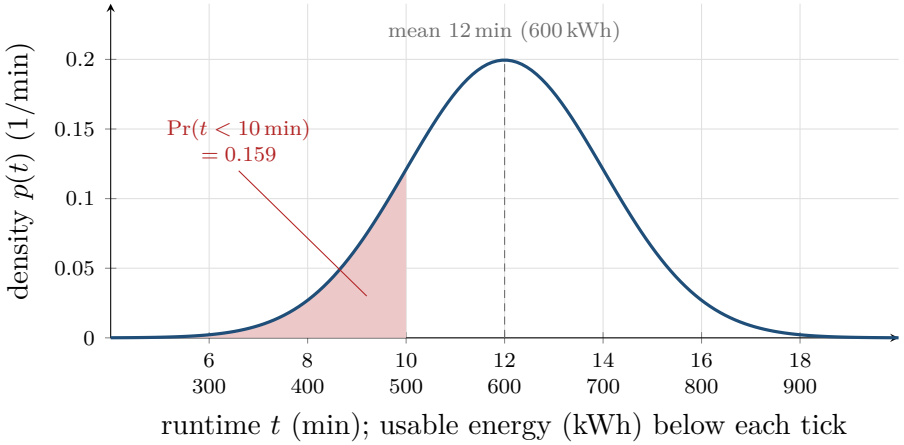


Figure 4.5: Runtime distribution for a usable energy of 600 ± 100 kWh at a battery power of 3 MW: $E_{\text{usable}} \sim \mathcal{N}(600, 100^2)$ kWh maps to $t = E_{\text{usable}}/P \sim \mathcal{N}(12, 2^2)$ min (equation (4.14)), illustrative. The shaded tail is the probability $\Phi(-1) = 0.159$ that the runtime is below 10 minutes.

4.5.4 The reliability question

The generator usually starts in 8 s, which against a ten-minute battery looks trivial. It is not trivial if the generator fails on its first attempt. The quantity that decides whether the critical load is interrupted is the probability that the battery runs out before the generator is available,

$$\Pr(T_{\text{gen}} > T_{\text{depl}}), \tag{4.15}$$

where T_{gen} is the time at which the generator can carry the load and T_{depl} the time at which the battery is depleted. Both are random variables. The UPS energy and the generator start are a race, and the reliability of the backup system is the probability that the generator wins it.

A worked example shows how the two uncertainties combine. Let the generator start on the first attempt with probability 0.98, at 8 s, which the battery always covers. With probability 0.02 the first attempt fails and recovery (a retry, a manual start or a transfer to another unit) takes a time R that is, illustratively, exponentially distributed with mean 10 minutes, so $\Pr(R > t) = e^{-t/10}$ for t in minutes. Let $T_{\text{depl}} \sim \mathcal{N}(12, 2^2)$ minutes, as in Figure 4.5. Then

$$\Pr(\text{interruption}) = 0.02 \times \mathbb{E}\left[e^{-T_{\text{depl}}/10}\right] = 0.02 \times e^{-12/10 + 2^2/(2\times 10^2)} = 0.02 \times e^{-1.18} = 0.0061, \tag{4.16}$$

using the normal moment-generating function $\mathbb{E}[e^{sT}] = e^{s\mu + s^2\sigma^2/2}$ with $s = -1/10$. About six outages in a thousand would interrupt the critical load. If the battery has aged to a mean runtime of 8 minutes, the same calculation gives $0.02 \times e^{-0.8+0.02} = 0.0092$, about one and a half times the risk. The generator's start probability and the battery's health enter the same number, and neither alone determines it.

4.5.5 Apparent versus real capacity

The state of charge is a fraction of the *present* usable capacity, not of the design capacity. A battery that reports SOC = 100 % at SOH = 65 % is at 100 % of 65 % of its nameplate. For

$E_{\text{nominal}} = 1\,\text{MWh}$,

$$E_{\text{actual}} = \text{SOC} \cdot \text{SOH} \cdot E_{\text{nominal}} = 1.00 \times 0.65 \times 1\,\text{MWh} = 0.65\,\text{MWh}.$$

The display shows a full battery, and the battery no longer meets its runtime requirement: at 5.21 MW it lasts 650/5208 h = 7.5 minutes, against the ten minutes it was installed for.

Figure 4.6 draws the distinction. Each column is one battery with a nameplate of 1 MWh, drawn as the dashed outline; the filled part is the energy it actually holds when its display reads SOC = 100 %. The vertical axis is energy in megawatt-hours. The three columns are a new battery, one at SOH = 0.8 and one at SOH = 0.65. The dashed red line is the energy needed to carry 5 MW for ten minutes through an inverter of efficiency 0.96, 0.868 MWh; the runtime at that load is printed under each column. All three report the same state of charge; only the first meets the requirement.

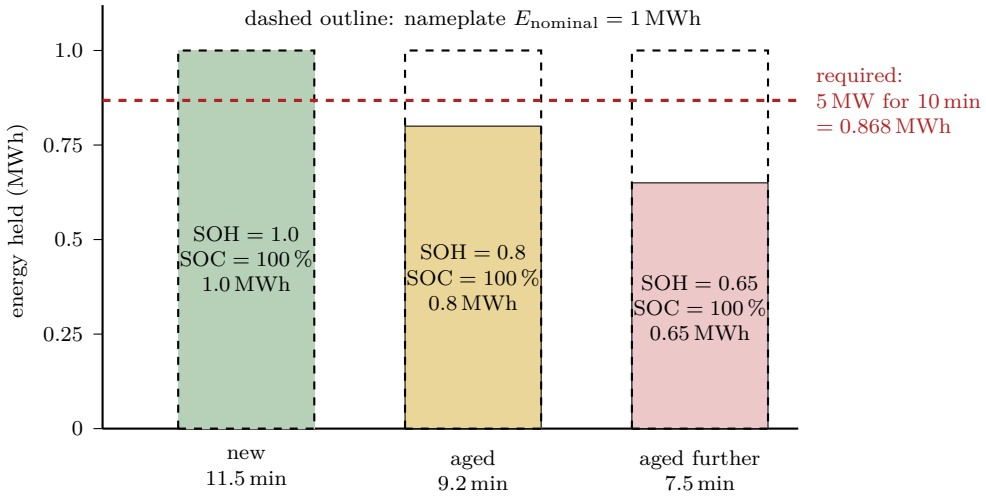


Figure 4.6: State of charge against state of health. Three batteries of 1 MWh nameplate, each displaying SOC = 100 %, hold SOH × 1 MWh: 1.0, 0.8 and 0.65 MWh. Against the 0.868 MWh needed for 5 MW over ten minutes with $\eta_{\text{inv}} = 0.96$, only the new battery is sufficient; runtimes at that load are printed below each column. Illustrative.

Principle 4.5. *A state of charge is a fraction of the present usable capacity. A full state of charge says nothing about whether the battery meets its runtime requirement unless the state of health is known as well.*

4.6 Dependence, diagnosis and dual use

The backup chain has several components whose failures are not independent, whose condition is mostly not measured, and whose stored energy has other uses. This section treats the dependence among generators, the diagnosis of a slow start, the split between what is measured, calculated and inferred, the use of the battery for grid services, and the posterior formulation that ties these together.

4.6.1 Multiple generators and dependence

Consider three generators, G1, G2 and G3, of which two are required to carry the load: an N+1 plant with $N = 2$. Each starts with probability $p = 0.98$. If the three starts were independent, the probability that at least two start would be

$$\Pr(\geq 2 \text{ start}) = p^3 + 3p^2(1-p) = 0.941192 + 3 \times 0.9604 \times 0.02 = 0.998816, \tag{4.17}$$

so the plant would fail to provide two units with probability 0.0012, about one outage in eight hundred. Independence is the assumption that makes the joint probability of the three outcomes the product of the individual ones, $\Pr(G_1, G_2, G_3) = \Pr(G_1)\ \Pr(G_2)\ \Pr(G_3)$.

The generators of a real plant share things. They may share a starting-air system, fuel from one tank (and so any contamination of it), a control network that issues the start signals, the switchgear through which they connect, and the environment of one site: the same cold night, the same flood, the same heat. Each shared element is a cause that can affect all three at once, and in its presence

$$\Pr(G_1, G_2, G_3) \neq \Pr(G_1)\ \Pr(G_2)\ \Pr(G_3). \tag{4.18}$$

The probability of the joint outcome has to be computed from a model of the shared causes, not from a product.

4.6.2 Common-cause failures

The plainest case is a blocked common fuel line. Three healthy generators, each mechanically sound, with starting batteries charged and controls working, all fail together because none receives fuel. The failure is not three failures; it is one, with three consequences.

Figure 4.7 draws the structure as a graph. Each box is a variable and each arrow a direct dependence. At the top is the state of the common fuel line, H_{fuel}, with the other shared causes of Section 4.6.1 in a dashed box beside it; they act on all three generators in the same way, and their arrows are left out to keep the drawing readable. In the middle row are the three generator outcomes, each with its own independent causes (its own starter, starting battery and governor) entering from below. At the bottom is the plant outcome, at least two of three started, and the critical load it supplies. A parent cause that points to several children creates dependence among them: once the state of the fuel line is unknown, learning that G1 failed raises the probability that G2 and G3 fail too, because one explanation of G1's failure is a blocked line that affects them all.

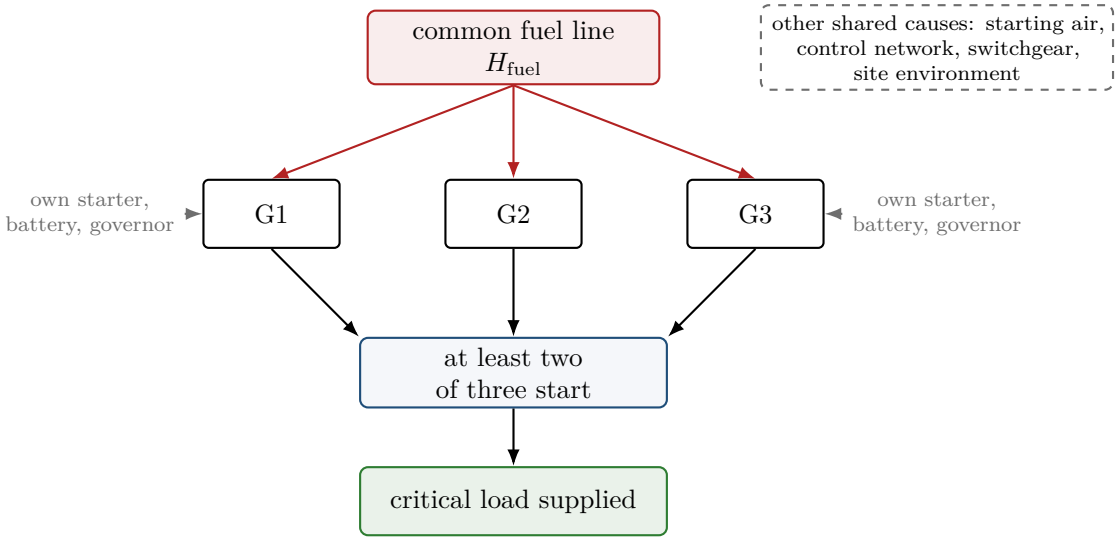


Figure 4.7: The common-cause structure of a three-generator plant. The common fuel line is a parent of all three generator outcomes, and other shared causes (dashed box, arrows omitted) act the same way; each generator also has its own independent causes. A shared parent makes the three outcomes dependent, so their joint probability is not the product of the individual ones.

The size of the effect follows from a simple model. Let the common cause be present with probability q, in which case all three generators fail; otherwise each starts independently

with probability p'. Hold each generator's own start probability at the observed 0.98, so that $(1-q)\,p' = 0.98$ and $p' = 0.98/(1-q)$. Then

$$\Pr(\geq 2 \text{ start}) = (1-q)\,[p'^3 + 3p'^2(1-p')] = (1-q)(3p'^2 - 2p'^3). \tag{4.19}$$

For $q = 0.01$, $p' = 0.98/0.99 = 0.9899$, and

$$\Pr(\geq 2 \text{ start}) = 0.99 \times (3 \times 0.97990 - 2 \times 0.97000) = 0.99 \times 0.99969 = 0.98970.$$

The probability that the plant fails to provide two units rises from 0.0012 to 0.0103, about nine times, although every generator still starts 98 % of the time when examined alone. The probability that all three fail rises from $0.02^3 = 8 \times 10^{-6}$ to about 0.01, more than a thousand times.

Figure 4.8 plots equation (4.19). The horizontal axis is the probability q of the common cause; the vertical axis is the probability that at least two of three generators start. The dashed gray line is the independent value 0.998816 of equation (4.17); the solid red curve is the value with a common cause. The two agree at $q = 0$ and separate at once: the plant's probability of success falls almost one for one with q, because each occurrence of the common cause defeats the redundancy entirely. At $q = 0.02$ the whole 2 % failure rate of each generator is common cause, $p' = 1$, and the plant is exactly as reliable as one generator, 0.98; the N+1 design adds nothing. Naive independence underestimates the risk, and by a factor that is set by the part of the failure probability that is shared, not by the individual failure probabilities.

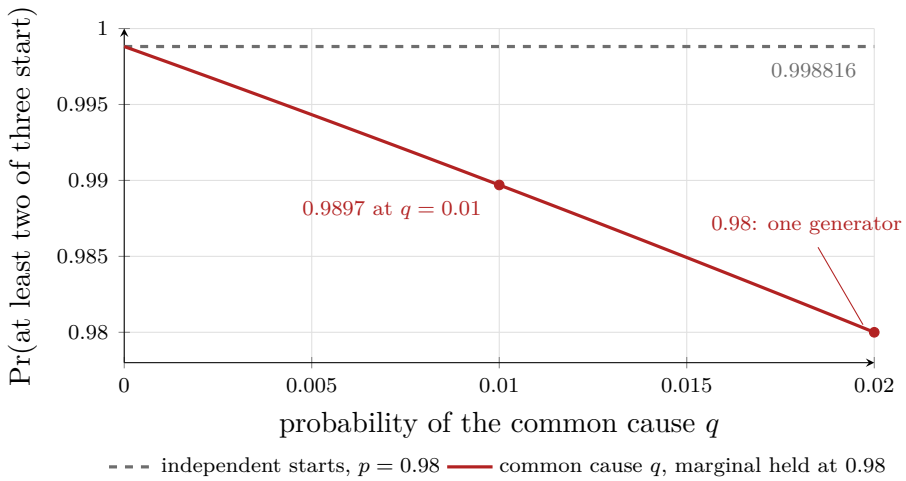


Figure 4.8: Probability that at least two of three generators start, independent versus with a common cause, illustrative. Independent: $3p^2 - 2p^3$ with $p = 0.98$. Common cause present with probability q, failing all three; otherwise independent starts with $p' = 0.98/(1-q)$: $(1-q)(3p'^2 - 2p'^3)$ (equation (4.19)). At $q = 0.02$ the plant is no more reliable than one generator.

Principle 4.6. *Redundancy does not imply independence. A shared parent cause, such as one fuel line, makes the failures of redundant units dependent, and a reliability computed as a product of individual probabilities underestimates the risk by a factor set by the shared part of the failure probability.*

4.6.3 Diagnosing a slow start

Backup power is not only a system to size; it is a system to diagnose. Generators are tested, typically monthly, by starting them and often running them under load. Suppose a test shows a start time well above the usual eight seconds. At least five explanations are possible:

- H_1: a weak starting battery, which cranks the engine slowly;
- H_2: a fuel problem, such as a partly blocked filter or air in the line, which lets the engine crank normally but delays firing;
- H_3: a temperature effect, a cold engine and cold starting battery on a cold day, which slows cranking and firing;
- H_4: degradation of the starter motor, which draws more current and turns the engine more slowly;
- H_5: a sensor or timing error, in which the start was normal and the recorded time is wrong.

A conventional monitoring system turns the test into an alarm, "start time exceeded threshold", and stops there. The alarm is true under all five hypotheses and distinguishes none of them. A richer approach asks for the probability of each hypothesis given the test data y,

$$\Pr(H_i \mid y) = \frac{\Pr(y \mid H_i)\,\Pr(H_i)}{\sum_j \Pr(y \mid H_j)\,\Pr(H_j)}, \tag{4.20}$$

which is Bayes' rule: the probability of a cause after the evidence is proportional to how likely that cause made the evidence, times how likely the cause was beforehand.

Table 4.3 works one illustrative case. The test data y are: a start time of 14 s, a low cranking speed, a dip of the starting-battery voltage deeper than in earlier tests, and a coolant temperature of 5°C. Each row is one hypothesis; the third column says what the hypothesis predicts in those data; the fourth gives an illustrative likelihood $\Pr(y \mid H_i)$; the last gives the posterior from equation (4.20) with equal prior probabilities of 0.2. The products $\Pr(y \mid H_i)\Pr(H_i)$ are 0.060, 0.004, 0.020, 0.030 and 0.002, summing to 0.116, and each posterior is its product divided by that sum.

Table 4.3: Hypotheses for a slow generator start, with illustrative likelihoods for the observed test data and the resulting posterior probabilities (equal priors of 0.2).

	Cause	What it predicts in the test data	$\Pr(y \mid H_i)$	$\Pr(H_i \mid y)$
H_1	weak starting battery	low cranking speed; deep battery voltage dip under cranking current	0.30	0.52
H_2	fuel problem	normal cranking speed and voltage dip; late firing	0.02	0.03
H_3	temperature effect	slower cranking on a cold day; moderate voltage dip	0.10	0.17
H_4	starter degradation	high cranking current, low speed; voltage dip from the high current	0.15	0.26
H_5	sensor or timing error	normal cranking speed and voltage; only the recorded time is long	0.01	0.02

The result is a ranking, not a verdict: a weak starting battery at 0.52, starter degradation at 0.26, the cold at 0.17. It also points to the next test. H_1 and H_4 both predict a deep voltage dip; they differ in where the problem sits, and a load test of the starting battery alone separates them. That test is worth more than another start, which would reproduce the same ambiguous evidence.

4.6.4 Measured, calculated and inferred

The backup chain is instrumented. A UPS typically reports the battery voltage V_{bat}, the battery current I_{bat}, the battery temperature T_{bat}, a state of charge, the UPS output power P_{UPS} and the DC bus voltage V_{DC}. A generator reports its output voltage, current, frequency and power, its engine speed in revolutions per minute, its coolant temperature T_{coolant}, its oil pressure p_{oil} and its fuel level or flow F_{fuel}. None of these is the quantity that decides whether the next outage is survived. They are evidence about hidden states.

The quantities that decide it are mostly not measured: the battery's state of health, its usable energy, the probability that a generator will start, the true availability of the fuel (as opposed to the level in the tank), and the reliability of the transfer switch. The book's three-way distinction applies directly. Table 4.4 sorts the quantities of this chapter into the three categories. The rows are measured, calculated and inferred; the columns are the UPS and battery, and the generator and transfer.

Table 4.4: Measured, calculated and inferred quantities of the backup power chain.

	UPS and battery	Generator and transfer
Measured	V_{bat}, I_{bat}, T_{bat}, V_{DC}, P_{UPS}	voltage, current, frequency, power, engine speed, T_{coolant}, p_{oil}, F_{fuel}, start time in tests
Calculated	$P = VI$; $E = \int P\,\mathrm{d}t$; a reported SOC by counting energy against an assumed capacity	energy delivered; fuel burn rate; $t_{\mathrm{fuel}} = F_{\mathrm{stored}}/\dot{F}$
Inferred	SOH; E_{usable}; remaining reliable runtime	probability of a successful start; true fuel availability; transfer-switch reliability; cause of a slow start

The reported state of charge sits in the calculated row for a reason: it is computed by the UPS from measured current or power and an assumed capacity, and if the assumed capacity is the nameplate rather than the present usable capacity, it is the misleading full display of Figure 4.6. Measured quantities are voltages, currents and temperatures; calculated ones follow from them by known relations such as $P = VI$ and $E = \int P\,\mathrm{d}t$; inferred ones, the state of health, a failure probability and the remaining reliable runtime, need a model and carry uncertainty.

4.6.5 Grid services against backup reserve

A large UPS battery is idle almost all the time, and an idle battery could earn money. It could shave the facility's peak demand, respond to demand-response calls, provide frequency response to the grid operator or buy energy when it is cheap and sell it when it is dear (energy arbitrage). Each use discharges the battery, and every kilowatt-hour used for a grid service is a kilowatt-hour not held for backup. At SOC = 80 % the energy available for backup is $E_{\mathrm{backup}} = 0.8\,E_{\mathrm{usable}}$, and with it the runtime falls by a fifth.

The conflict is resolved by a constraint on risk rather than on energy:

$$\Pr(t_{\mathrm{runtime}} < t_{\mathrm{required}}) < \epsilon, \tag{4.21}$$

where ϵ is the acceptable probability of falling short. With the battery of Section 4.5.3, $E_{\mathrm{usable}} \sim \mathcal{N}(600, 100^2)$ kWh at 3 MW, and a required runtime of 8 minutes, which needs 400 kWh:

- at SOC = 100 %, $\Pr(t < 8\,\mathrm{min}) = \Phi(-2) = 0.023$;
- at SOC = 80 %, the available energy is $\mathcal{N}(480, 80^2)$ kWh, the runtime $\mathcal{N}(9.6, 1.6^2)$ minutes, and $\Pr(t < 8\,\mathrm{min}) = \Phi(-1) = 0.159$.

For $\epsilon = 0.05$ the lowest admissible state of charge solves $\mathrm{SOC} \times (600 - 1.645 \times 100) \geq 400$, where 1.645 is the standard normal quantile at 0.95:

$$\mathrm{SOC} \geq \frac{400}{435.5} = 0.92.$$

Only the top 8 % of the usable energy can be offered for grid services without violating the constraint, and less as the battery ages or the uncertainty about its usable energy grows. Because the state of charge, the state of health and the load all change in time, holding equation (4.21) while earning grid-service revenue is a stochastic control and optimization problem: each dispatch decision must keep the probability of a backup shortfall below ϵ over the period in which an outage could arrive.

4.6.6 A posterior formulation

The pieces of this section assemble into one formulation. Collect the hidden states of the backup chain in a vector

$$x = \left[\mathrm{SOC},\ \mathrm{SOH},\ E_{\mathrm{usable}},\ G_1,\ G_2,\ F_{\mathrm{fuel}},\ \ldots\right]^{\mathsf{T}}, \tag{4.22}$$

where G_1, G_2 are the start readiness of the generators and F_{fuel} is the usable fuel, and the measurements in a vector

$$y = \left[V_{\mathrm{bat}},\ I_{\mathrm{bat}},\ T_{\mathrm{bat}},\ P_{\mathrm{UPS}},\ \text{generator test data},\ \ldots\right]^{\mathsf{T}}. \tag{4.23}$$

The object of interest is the posterior distribution of the states given the measurements,

$$p(x \mid y) \propto p(y \mid x)\, p(x), \tag{4.24}$$

where $p(y \mid x)$ is the measurement model (how a battery with a given state of health responds in voltage to a given current; how a generator with a weak starting battery shows up in a test) and $p(x)$ is the prior (aging models, maintenance history, fleet statistics). From it follows the probability that matters to the operator,

$$\Pr(\text{critical load interrupted} \mid y) = \int \Pr(\text{interrupted} \mid x)\, p(x \mid y)\, \mathrm{d}x, \tag{4.25}$$

where $\Pr(\text{interrupted} \mid x)$ is the race of equation (4.15), evaluated with the common-cause structure of Section 4.6.2 for a given state. A dashboard that shows "UPS status = green" reports that no alarm is active. Equation (4.25) reports how likely the next outage is to reach the IT load, given everything measured so far, and it changes when a test result, a temperature history or a discharge record arrives.

4.6.7 Backup power as a dynamic reliability problem

The conceptual distinction of the chapter is between two jobs. The battery's job is instantaneous continuity: there is no gap at $t = 0$. The generators' job is longer-duration supply: hours or days on stored fuel. The chain grid $\rightarrow$ UPS and battery $\rightarrow$ generator is coordinated by sensors, controls and transfer logic, and its reliability depends on all of them: a perfect battery and perfect generators fail together if the transfer switch does not transfer.

The question this leads to is not whether the battery capacity exceeds the required energy. It is whether enough electrical capacity remains available throughout the disturbance. Written as an event, the critical load is interrupted if at some time during the outage the available supply falls below the critical demand,

$$\text{interruption} = \Big\{\, \exists\, t \in [0, T_{\mathrm{out}}] : P_{\mathrm{gen,avail}}(t) + P_{\mathrm{bat,avail}}(t) < P_{\mathrm{crit}}(t) \,\Big\}, \tag{4.26}$$

where T_{out} is the outage duration, the generator term depends on which units started and on the fuel, and the battery term falls to zero when its energy is exhausted. The probability of this event depends jointly on the state of charge, the state of health, the load, the inverter efficiency, the generator start time, the availability of each generator, the fuel, the switchgear and the controls. Every one of these was treated separately in this chapter; the event couples them in time. Backup power is a dynamic reliability problem, not a battery-sizing problem.

> **Principle 4.7.** *The critical question for backup power is not whether battery capacity exceeds the required energy but whether enough electrical capacity remains available throughout the disturbance; its answer is a probability that depends jointly on the battery, the load, the generators, the fuel and the switching.*

4.7 What to retain

Definition 4.1 (Backup power as a dynamic reliability problem)**.** The UPS battery provides instantaneous continuity and the generators provide duration; whether the critical load stays powered is the probability that available electrical capacity, from the battery while its energy lasts and from the generators once they start, stays above the critical demand throughout the disturbance, a probability that depends jointly on state of charge, state of health, load, efficiency, generator start, generator availability, fuel, switchgear and controls.

Table 4.5 lists the chapter's objects and the subsection that defines each.

4.8 Exercises

Exercise 4.1. A UPS must carry a 3 MW critical load for 8 minutes through an inverter of efficiency 0.95. (a) Compute the ideal energy and the energy the battery must deliver at its terminals. (b) Apply a battery internal efficiency of 0.97, an end-of-life state of health of 0.8, an allowable depth of discharge of 0.9 and a reserve margin of 1.10, as in Table 4.1, and give the installed nominal capacity. (c) What is the battery power at full load, and what is the ratio of that power to the installed capacity, in units of per hour? (d) How much heat does the inverter release into the UPS room during the discharge, in kilowatts?

Exercise 4.2. A facility has a critical load of 18 MW and generators of 2.5 MW. (a) How many generators does it need, and how many does an N+1 design install? (b) The plant burns 0.26 liters of diesel per kilowatt-hour and stores 150 000 liters; how long does the fuel last at 18 MW? (c) During an outage two generators fail to start. The total load is 21 MW, split as 13 MW of critical IT, 5 MW of its cooling, 2 MW of less critical computing and 1 MW of building services. How much load must be shed, and from which tiers? (d) The operator proposes instead to take the whole shortfall from the cooling of the critical IT. If the heat the cooling removes falls in proportion to its electrical power, what is the heat imbalance $\dot{Q}_{\mathrm{IT}} - \dot{Q}_{\mathrm{cool}}$ in the data halls? With an effective thermal capacitance of 400 MJ/K, how long do the halls take to warm by 5 K?

Exercise 4.3. A battery's usable energy is estimated at 900 ± 150 kWh (normal, mean and standard deviation), and it carries a 4 MW load through an inverter of efficiency 0.96. (a) Give the mean and standard deviation of the runtime in minutes. (b) Compute the probability that the runtime is below 10 minutes. (c) Show that a requirement $\Pr(t < 10\,\mathrm{min}) < 0.05$ cannot be met even at $\mathrm{SOC} = 100\,\%$. (d) For a requirement of 8 minutes at the same $\epsilon = 0.05$, find the lowest state of charge at which the battery may be held while it provides grid services.

Table 4.5: The objects of the backup power chain and where each is defined.

Object	Expression	Role	Defined in
Outage timeline	utility, battery, generator	order and speed of the sources	4.1.1
Double conversion	equations (4.1), (4.2)	battery always on the path	4.1.2
Energy requirement	$E = P_{\text{load}} t$	ideal battery energy	4.2.1
Power and energy	$E = Pt$	two separate requirements	4.2.2
State of charge	$\mathrm{dSOC}/\mathrm{d}t = -P_{\text{bat}}/E_{\text{usable}}$	state of the battery	4.2.3
Inverter loss	$P_{\text{bat}} = P_{\text{load}}/\eta_{\text{inv}}$	extra power, extra heat	4.2.4
Runtime window	5 to 15 minutes	time to recover a failed start	4.3.2
Generator count	$N = \lceil P_{\text{crit}}/P_{\text{unit}} \rceil$, N+1	capacity with one unit missing	4.3.3
Fuel runtime	$t_{\text{fuel}} = F_{\text{stored}}/\dot{F}$	long-duration limit	4.3.4
Transfer switches	ATS, STS	source selection	4.3.5, 4.3.6
Load tiers	tiers 1 to 4	shedding order	4.4.1
Thermal buffer	$C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = \dot{Q}_{\text{IT}} - \dot{Q}_{\text{cool}}$	time before overheating	4.4.3
Usable energy	$E_{\text{usable}} = \text{SOH} \cdot E_{\text{nominal}}$	capacity after aging	4.5.2
Runtime distribution	$t = E_{\text{usable}}/P$, $p(t)$	runtime under uncertainty	4.5.3
Reliability race	$\Pr(T_{\text{gen}} > T_{\text{depl}})$	interruption probability	4.5.4
Common cause	equation (4.19)	dependence among generators	4.6.2
Diagnosis	$\Pr(H_i \mid y)$	ranked causes of a symptom	4.6.3
Resilience constraint	$\Pr(t_{\text{runtime}} < t_{\text{required}}) < \epsilon$	backup reserve against grid services	4.6.5
Posterior	$p(x \mid y)$, equation (4.25)	risk given the measurements	4.6.6

Exercise 4.4. Three generators, two required, each start with probability 0.97. (a) Compute the probability that at least two start if the starts are independent. (b) A shared starting-air system fails with probability $q = 0.01$ and prevents all three from starting; holding each generator's start probability at 0.97, compute p' and the probability that at least two start, and the ratio of the two failure probabilities. (c) A monthly test of one generator shows a slow start with a normal cranking speed, a normal starting-battery voltage dip and late firing. Using Table 4.3 as a guide, which hypothesis does this evidence favor, and is it a cause that the other two generators might share?

4.9 Connection to the TERRA framework

The backup chain is an electrical conservation network with storage, switching and uncertain components, and it maps onto the framework's graph directly.

- **Sources, buses and converters are nodes and edges.** The utility connection, the generator bus, the UPS input, the DC bus and the critical bus are nodes of the electrical network, each with a power balance. The transfer switch, rectifier, inverter and static switch are edges whose flows carry power, with efficiencies such as η_{inv} in equation (4.8); the loss on each converter edge is a source of heat on a node of the thermal network, the UPS room, as in Figure 4.2.

- **The battery is a storage component with a state and bounds.** Equation (4.5) is the balance of a storage node whose state is the energy E, bounded by $E_{\text{usable}} = \text{SOH} \cdot E_{\text{nominal}}$, with a power bound separate from the energy bound (Section 4.2.2). Forward simulation in time produces Figure 4.3; the switching of the outage timeline is a change of which edges are active.

- **The thermal buffer is a coupled node.** Equation (4.11) is the energy balance of a thermal node whose capacitance depends on which air and water masses the fans and pumps keep connected, so a load-shedding decision on the electrical network changes the topology of the thermal network.

- **Hidden states are inferred, not displayed.** Voltages, currents and temperatures are measured; power and energy are calculated; SOH, E_{usable} and the start readiness of each generator are inferred as the posterior $p(x \mid y)$ of equation (4.24), with the common fuel line of Figure 4.7 represented as a shared parent of the generator components rather than as independent failure rates.

- **Risk and diagnosis are posterior quantities.** The interruption probability of equation (4.25) is a risk stated as the probability that available capacity falls below required capacity; the slow-start diagnosis of Table 4.3 is a ranked posterior over hypotheses for one symptom, and the starting-battery load test is the next measurement chosen for the information it adds.

In summary, the UPS battery gives the critical load continuity because it sits on the power path, the generators give it duration once they start, and the transfer between them is a race between the battery's energy and the generator's start whose outcome depends on aging, dependence and switching that the measurements reveal only indirectly; the reliability of backup power is therefore a probability conditioned on those measurements, not a ratio of capacities. The next chapter, Chapter 5, follows the electrical power that the chain delivers into the heat that the IT equipment releases and the balance that the cooling system must close.

Chapter 5

Heat generation and thermal balance: generation, transport and storage

5.1 Conservation of energy

The previous chapters followed electrical power from the utility connection through transformers, uninterruptible power supplies and distribution units to the IT equipment. That power does not stay electrical. Almost all of it becomes heat inside the building, and the facility exists as much to remove that heat as to deliver the power. The question of this chapter is

how electrical power becomes heat in a data center, and how that heat is accounted for as it moves from the chips to the outdoor environment.

Nearly every cooling, capacity and diagnostic question in the rest of the book returns to a heat balance: whether a cooling plant is large enough, how long a room survives the loss of a chiller, why a rack inlet is warm, whether a chilled-water meter can be trusted. This chapter builds the thermal model explicitly, from the balance of one control volume to a network of coupled volumes, and ends with the balance written as an inference problem: given the measurements, which states, parameters and causes are consistent with them.

Jargon: control volume, thermal capacitance, thermal resistance, sensible heat, recirculation, bypass, time constant

A *control volume* is a region of space chosen for accounting, with a boundary across which energy and mass flows are counted: a server, a rack, an aisle, a room or a water loop. *Thermal capacitance* C_{th}, in joules per kelvin, is the energy needed to raise the temperature of a body by one kelvin; for a mass m of specific heat c_p it is mc_p. *Thermal resistance* R_{th}, in kelvin per watt, is the temperature difference needed to drive one watt of heat along a path; its inverse is a thermal conductance. *Sensible heat* is heat that changes the temperature of a substance without changing its phase, as opposed to *latent heat*, which evaporates or condenses water without a change of temperature; the air and water balances of this chapter are sensible. *Recirculation* is hot exhaust air that flows back to the inlet of IT equipment instead of returning to the cooling unit. *Bypass* is cold supply air that returns to the cooling unit without passing through IT equipment. The *time constant* $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$, in seconds, is the time a temperature takes to cover about 63 % of a step toward a new equilibrium.

Key idea

Every temperature in a data center is the state of an energy balance with three parts: servers generate heat, air and water transport it, and thermal masses store it. A temperature that is too high says that generation is too high, transport is too weak, storage is changing, or the measurement is wrong; the balance is the tool for deciding which.

5.1.1 The energy balance of a control volume

Energy is conserved. For any control volume, the rate at which the energy stored inside it changes equals the rate at which heat enters, minus the rate at which heat leaves, plus the rate at which energy is converted into heat inside the volume:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}} + \dot{Q}_{\text{gen}}. \tag{5.1}$$

Every term is a power, in watts. E is the internal energy of whatever is inside the boundary (air, metal, water, silicon); $\dot{Q}_{\text{in}}$ and $\dot{Q}_{\text{out}}$ are heat flows across the boundary, carried by air or water streams or conducted through walls; and $\dot{Q}_{\text{gen}}$ is internal generation, the electrical power dissipated inside the volume. Read in words,

$$\text{rate of accumulation} = \text{heat in} - \text{heat out} + \text{internal generation}.$$

For a body at one uniform temperature, without a change of phase and with a specific heat that does not vary much over the temperature range, the stored energy is proportional to temperature, $E \approx C\,T$ measured from any convenient reference, and only its changes matter. Substituting into equation (5.1) gives the form used throughout the book,

$$C\,\frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}} + \dot{Q}_{\text{gen}}, \tag{5.2}$$

where C is the thermal capacitance of the volume. The equation turns a statement about energy into a statement about temperature: when the right-hand side is positive the volume warms, when it is negative the volume cools, and when it is zero the temperature holds. The last case is the steady state, $\dot{Q}_{\text{out}} = \dot{Q}_{\text{in}} + \dot{Q}_{\text{gen}}$: all the heat that arrives or is generated also leaves.

Figure 5.1 draws the balance. The dashed outline is the boundary of the control volume. The circle inside is the stored energy $E \approx CT$. The arrow entering from the left is heat arriving from warmer surroundings, the arrow leaving on the right is heat carried away by the cooling system, and the arrow entering from above is electrical power, which crosses the boundary as electricity and becomes heat inside it. The balance below the volume is equation (5.2). What the figure makes explicit is that temperature is not an input: it is the result of the three flows, integrated over time by the capacitance.

Principle 5.1. *The temperature of a control volume changes only when the heat entering plus the heat generated differs from the heat leaving. At steady state the three balance and the stored energy is constant.*

5.1.2 Server heat

Apply the balance to one server. A server that draws $P_{\text{server}} = 800\,\text{W}$ of electrical power releases, in steady operation, about $\dot{Q}_{\text{server}} \approx 800\,\text{W}$ of heat into the air that its fans move through it. A rack of fifty such servers draws $P_{\text{rack}} = 50 \times 0.8 = 40\,\text{kW}$ and releases $\dot{Q}_{\text{rack}} \approx 40\,\text{kW}$. At every scale the electrical IT load and the thermal IT load are the same number,

$$\dot{Q}_{\text{IT}} \approx P_{\text{IT}}, \tag{5.3}$$

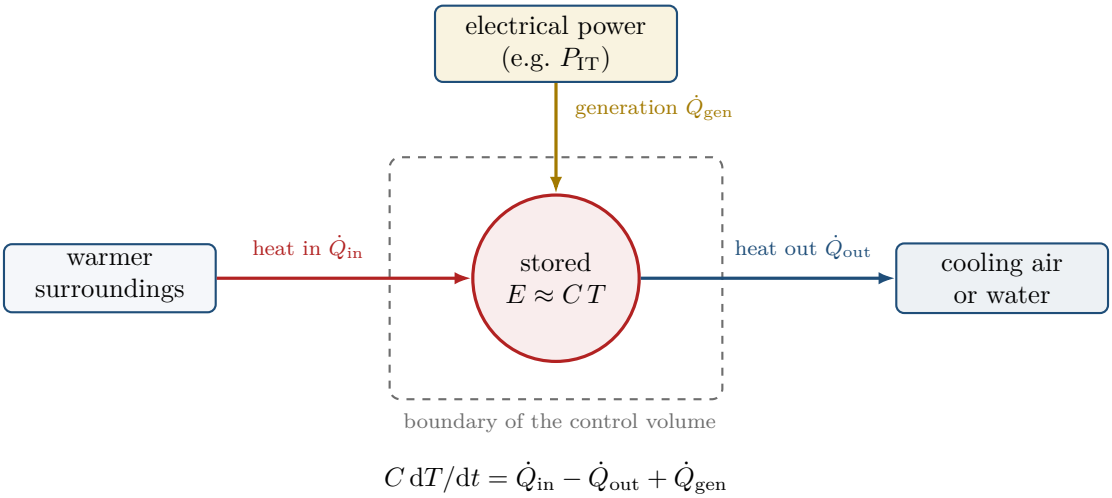


Figure 5.1: The energy balance of a control volume, equation (5.2). Heat enters and leaves across the boundary, electrical power is converted into heat inside it, and the difference accumulates as stored energy and appears as a change in temperature.

which is the most useful identity in data center thermal engineering. It means that the heat a cooling system must remove can be read from the electrical meters.

The identity also fixes the status of the two quantities in the book's recurring distinction. Rack power P_{rack} is *measured*, by the metered power distribution units of Chapter 3; rack heat $\dot{Q}_{\text{rack}}$ is *calculated* from it by equation (5.3). Because electrical meters are usually more accurate than flow meters and temperature-difference measurements, the electrically calculated heat is often the most reliable heat number a facility has, and later sections use it as the reference against which air-side and water-side estimates are checked. The identity holds on average over the time it takes heat to leave the server: when the load steps up, the heat sinks and the chassis absorb part of the extra energy for seconds to a minute before the air carries it out, so the instantaneous heat release lags the electrical power by the server's own small thermal capacitance.

5.1.3 Why computation becomes heat

A processor performs computation by changing the state of transistors and memory cells. Each change moves electric charge onto or off a small capacitance through a resistive channel, and the energy of that charge is dissipated in the channel: a switching event of a node of capacitance C_{node} at supply voltage V dissipates on the order of $C_{\text{node}}V^2$ per cycle. Leakage currents flow through transistors even when they do not switch, resistive losses heat every conductor, and the power supply units and voltage regulators inside the server convert power with losses of their own. All of these are dissipative processes, and all of them end as heat.

The information produced by the computation carries negligible energy. The thermodynamic minimum energy to erase one bit is $kT\ln 2$, about 2.9×10^{-21} J at 300 K, where k is Boltzmann's constant; even 10^{18} bit erasures per second correspond to about 3 mW, against 800 W of electrical input. The other paths by which energy could leave without becoming heat in the room are also small: fan power first becomes kinetic energy of the air and then heat through friction, light from indicator lamps is absorbed by the walls, and the optical and electrical signals that leave the building on network cables carry milliwatts. To within a fraction of a percent,

$P_{\text{IT}} \longrightarrow \dot{Q}_{\text{IT}}$: every watt of electricity delivered to the IT equipment becomes a watt of heat.

The same is true of the losses of the power chain, transformers and uninterruptible power

supplies, and of lighting. A data center's heat balance is therefore its electrical balance read from the other end, and Section 5.3.1 uses that fact to set the load on the cooling plant.

5.2 The rack and the air stream

This section writes the balance for one rack, introduces the equation by which air carries heat away, shows that the air temperature rise is fixed by power and airflow, and identifies the first diagnostic ambiguity that a single temperature leaves open.

5.2.1 The rack heat balance

Take one rack as the control volume. Its generation is its electrical power P_{rack}; the heat that leaves is whatever the air stream carries out, $\dot{Q}_{\text{removed}}$. With C_r the thermal capacitance of the rack (servers, chassis, the air inside) and T_r a representative rack temperature, equation (5.2) becomes

$$C_r \frac{\mathrm{d}T_r}{\mathrm{d}t} = P_{\text{rack}} - \dot{Q}_{\text{removed}}. \tag{5.4}$$

Three cases follow from the sign of the right-hand side:

- $P_{\text{rack}} = \dot{Q}_{\text{removed}}$: the rack is in steady state and its temperature holds;
- $P_{\text{rack}} > \dot{Q}_{\text{removed}}$: energy accumulates and the rack temperature rises;
- $P_{\text{rack}} < \dot{Q}_{\text{removed}}$: the rack gives up stored energy and cools, as it does after a load drop.

The second case is the essence of a cooling shortfall. A shortfall is not a temperature; it is a difference between generation and removal, and temperature is its integral over time. A shortfall of a few percent sustained for long enough raises any temperature without limit unless something makes removal grow. A temperature reading therefore says how much energy has accumulated, not how large the present imbalance is; the imbalance is read from the slope.

5.2.2 Heat removal by air

Air passing through a rack enters at the inlet temperature T_{in} and leaves at the outlet temperature T_{out}. Each kilogram of air that warms by one kelvin carries away $c_{p,\text{air}}$ joules, so a mass flow $\dot{m}_{\text{air}}$ carries

$$\dot{Q}_{\text{air}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, (T_{\text{out}} - T_{\text{in}}). \tag{5.5}$$

This is a sensible-heat balance on the air stream: the heat it removes from the rack equals the heat it gains as it warms. For air near room conditions $c_{p,\text{air}} \approx 1.0\,\text{kJ/(kg K)}$.

Take an illustrative rack with $\dot{m}_{\text{air}} = 3\,\text{kg/s}$ of air warming by $T_{\text{out}} - T_{\text{in}} = 12°\text{C}$. The air removes

$$\dot{Q}_{\text{air}} = 3\,\text{kg/s} \times 1.0\,\text{kJ/(kg K)} \times 12\,\text{K} = 36\,\text{kW}.$$

If the rack is the 40 kW rack of Section 5.1.2, then by equation (5.4) the rack accumulates $40 - 36 = 4\,\text{kW}$ and $\mathrm{d}T_r/\mathrm{d}t > 0$: it is warming.

The imbalance is a snapshot of a transient, and the balance says how it ends. As the rack warms, the air leaving it warms too; with the inlet temperature and the airflow unchanged, the removal grows with the outlet temperature until it reaches 40 kW, at a rise of $40/(3{\times}1.0) = 13.3\,\text{K}$. With $T_{\text{in}} = 20°\text{C}$, the outlet moves from 32°C toward 33.3°C and the rack settles. The air stream is self-correcting in this sense, and a dangerous shortfall is one where that correction cannot happen within limits: the components reach their maximum temperature before removal catches up, or the air delivered to the inlet is itself warming because the room as a whole is short of cooling.

The measurement status of the three quantities in equation (5.5) differs. Inlet and outlet temperatures are *measured* by sensors at the front and back of the rack. Airflow through a rack is rarely measured directly; it is *calculated* from fan speeds through a fan curve, or *inferred* from the balance itself, $\dot{m}_{\text{air}} = P_{\text{rack}}/(c_{p,\text{air}}(T_{\text{out}} - T_{\text{in}}))$, using the measured electrical power.

5.2.3 The air temperature rise

Solving equation (5.5) at steady state, with $\dot{Q}_{\text{air}} = \dot{Q}_{\text{rack}}$, gives the temperature rise across the rack,

$$\Delta T = T_{\text{out}} - T_{\text{in}} = \frac{\dot{Q}_{\text{rack}}}{\dot{m}_{\text{air}}\, c_{p,\text{air}}}. \tag{5.6}$$

The rise is not arbitrary: power and airflow fix it. For $\dot{Q}_{\text{rack}} = 40\,\text{kW}$ and $\dot{m}_{\text{air}} = 4\,\text{kg/s}$,

$$\Delta T = \frac{40\,\text{kW}}{4\,\text{kg/s} \times 1.0\,\text{kJ/(kg K)}} = 10\,\text{K};$$

with the airflow halved to 2 kg/s, the rise doubles to 20 K. The IT load is the same in both cases. A reduced airflow alone produces a high exhaust temperature, which is why a failed fan, a blocked filter or a cable bundle behind a server shows up first as a hot exhaust. In volume, with air at about $1.2\,\text{kg/m}^3$, 4 kg/s is about $3.3\,\text{m}^3/\text{s}$ of air for one rack.

Figure 5.2 plots equation (5.6). The horizontal axis is the rack airflow in kilograms per second; the vertical axis is the air temperature rise in kelvin. The solid curve is the 40 kW rack; the dashed curve is a rack at twice the power, 80 kW. The square marks the baseline operating point, 40 kW at 4 kg/s and a 10 K rise. The two dots mark the two ways of reaching a 20 K rise from it: halving the airflow at the same power, and doubling the power at the same airflow. Both curves are hyperbolas: at high airflow the rise is small and insensitive to further increases, at low airflow it climbs steeply, so the same loss of airflow matters more in a rack that is already short of air.

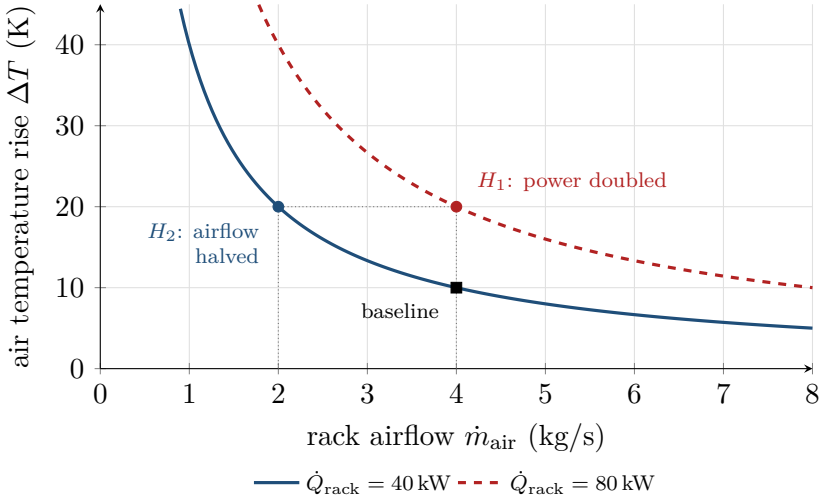


Figure 5.2: Air temperature rise across a rack against airflow, $\Delta T = \dot{Q}_{\text{rack}}/(\dot{m}_{\text{air}} c_{p,\text{air}})$ (equation (5.6)) with $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$, for 40 kW and 80 kW (illustrative). From the baseline, halving the airflow and doubling the power give the same 20 K rise.

5.2.4 The first diagnostic ambiguity

Suppose the outlet temperature of a rack rises while its inlet holds. Equation (5.6), written as $\Delta T = P_{\text{rack}}/(\dot{m}_{\text{air}} c_{p,\text{air}})$, offers two explanations:

- H_1: the rack power P_{rack} increased;
- H_2: the rack airflow $\dot{m}_{\text{air}}$ decreased.

Figure 5.2 shows them as the two dots at $\Delta T = 20\,\text{K}$. Both predict exactly the same temperature rise, because the temperature rise depends on power and airflow only through their ratio. One temperature measurement, or one temperature difference, cannot distinguish them, however precise it is. This is the book's first encounter with *identifiability*: a quantity is identifiable from a set of measurements when different values of it produce different predicted measurements. Here the ratio $P_{\text{rack}}/\dot{m}_{\text{air}}$ is identifiable from ΔT; neither factor is.

The ambiguity is resolved by adding a measurement that responds to one hypothesis and not the other. A rack power meter does that: under H_1 it reads 80 kW, under H_2 it still reads 40 kW. A fan-speed reading does it from the other side. Without either, the two hypotheses remain equally consistent with the data, and any statement about which one happened is an assumption rather than an inference. The pattern recurs through the chapter: a balance equation relates several quantities, a measurement constrains a combination of them, and diagnosis consists in finding the measurements that separate the combinations.

Principle 5.2. *A temperature rise across a rack determines only the ratio of power to airflow. Separating a load increase from an airflow loss requires a second, independent measurement: rack power or fan speed.*

5.3 The room, the water loop and liquid cooling

This section widens the control volume from the rack to the room, follows the heat through the cooling coil into the chilled water and on to the environment, and explains why water, and liquid brought to the chip, carries heat so much more compactly than air.

5.3.1 The room heat balance

Take the whole computer room as the control volume. Its generation includes the IT heat, the losses of the electrical equipment inside the room (power distribution units, uninterruptible power supplies and transformers when they are located there) and other sources such as lighting, people and heat entering through the building envelope. Its removal is the heat taken out by the cooling units. With C_{room} the effective thermal capacitance of the room and T_{room} a representative temperature,

$$C_{\text{room}} \frac{\mathrm{d}T_{\text{room}}}{\mathrm{d}t} = \dot{Q}_{\text{IT}} + \dot{Q}_{\text{elec losses}} + \dot{Q}_{\text{other}} - \dot{Q}_{\text{cooling}}, \tag{5.7}$$

and at steady state the cooling must remove everything that is generated,

$$\dot{Q}_{\text{cooling}} \approx \dot{Q}_{\text{IT}} + \dot{Q}_{\text{elec losses}} + \dot{Q}_{\text{other}}. \tag{5.8}$$

Every term on the right is electrical power, or is close to it, by the argument of Section 5.1.3. The cooling load is set by the electrical meters.

Example 5.1. A room has $P_{\text{IT}} = 5\,\text{MW}$, electrical losses of 0.3 MW in its power equipment, and 0.1 MW of lighting and other gains. By equation (5.8) the cooling must remove

$$\dot{Q}_{\text{cooling}} = 5.0 + 0.3 + 0.1 = 5.4\,\text{MW}.$$

If it removes only 5.1 MW, then by equation (5.7) 0.3 MW is stored and temperatures rise. With the illustrative effective capacitance of 500 MJ/K used in Section 5.4.4, the room warms at $0.3/500 = 6 \times 10^{-4}\,\text{K/s}$, about 2.2 K per hour. The fans of the cooling units themselves add heat to the air they move; in a full account their power is one more term of $\dot{Q}_{\text{other}}$ that the coils must remove.

5.3.2 The heat chain through the coil

Heat does not disappear at the cooling coil. A computer room air handler (CRAH), a unit that blows room air across a coil carrying chilled water, moves the 5.4 MW from the air into the water. Across the coil the air-side and water-side heat flows are equal, apart from small losses and the coil's own storage,

$$\dot{Q}_{\text{air}} \approx \dot{Q}_{\text{water}}. \tag{5.9}$$

The water carries the heat through pipes to a chiller, whose evaporator transfers it again, into a refrigerant; the refrigeration cycle raises its temperature with compressor work W_{comp}, and the condenser rejects heat to the environment through a cooling tower or a dry cooler. The chain is

$$\text{IT equipment} \rightarrow \text{air} \rightarrow \text{water} \rightarrow \text{refrigeration system} \rightarrow \text{environment},$$

and energy balances at every stage. At the condenser the balance includes the compressor work, $\dot{Q}_{\text{cond}} = \dot{Q}_{\text{evap}} + W_{\text{comp}}$, so the heat rejected outdoors exceeds the heat collected indoors. With an illustrative coefficient of performance $\text{COP} = \dot{Q}_{\text{evap}}/W_{\text{comp}} = 5$, the 5.4 MW collected needs $W_{\text{comp}} = 1.08$ MW and the condenser rejects 6.48 MW. Chapter 7 treats the equipment of each stage.

Figure 5.3 draws the chain from top to bottom. Each box is a medium that holds the heat for one stage; each downward arrow is a transfer between media, labeled on the right with the device that performs it and the balance that holds across it. The arrows entering from the left are electrical inputs, each of which adds heat to the chain: the IT power at the top and the compressor work at the refrigeration stage (fan and pump power add smaller amounts at the air and water stages). Reading down the right-hand column gives the redundancy that a balance provides: the same heat is estimated electrically at the top, from air temperatures and airflow at the coil, from water temperatures and flow at the loop, and from refrigerant data at the chiller. Each estimate is a check on the others.

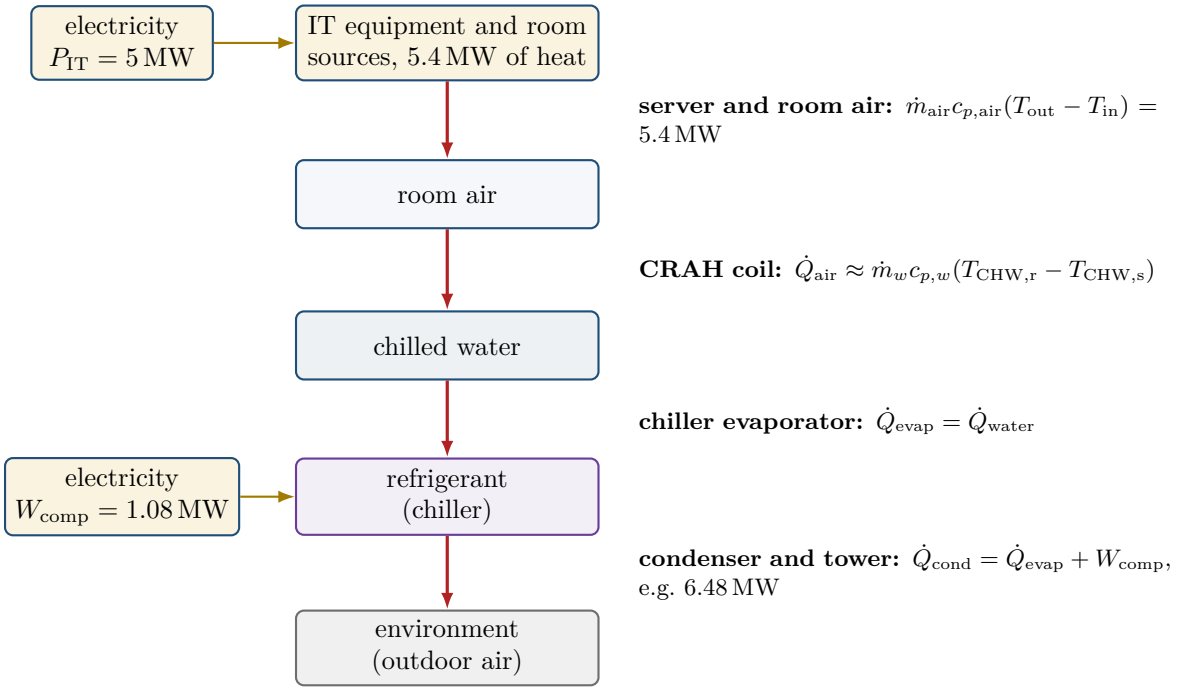


Figure 5.3: The heat chain of Example 5.1: heat generated by the IT equipment and the room's other sources passes from air to chilled water to refrigerant to the environment, with a balance at every stage. The compressor work (illustrative COP = 5) is added at the refrigeration stage, so the condenser rejects more heat than the room generates.

5.3.3 The chilled-water balance

The chilled-water loop obeys the same sensible-heat equation as the air stream, with the properties of water. Water leaves the plant at the chilled-water supply temperature $T_{\mathrm{CHW,s}}$, picks up heat at the coils, and returns at $T_{\mathrm{CHW,r}}$; with mass flow $\dot{m}_w$,

$$\dot{Q}_{\mathrm{water}} = \dot{m}_w \, c_{p,w} \, (T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}}), \qquad c_{p,w} \approx 4.18\,\mathrm{kJ/(kg\,K)}. \tag{5.10}$$

For an illustrative loop with $\dot{m}_w = 200\,\mathrm{kg/s}$ and a 5 K temperature difference,

$$\dot{Q}_{\mathrm{water}} = 200\,\mathrm{kg/s} \times 4.18\,\mathrm{kJ/(kg\,K)} \times 5\,\mathrm{K} = 4180\,\mathrm{kW} = 4.18\,\mathrm{MW}.$$

Carrying the 5.4 MW of Example 5.1 at the same 5 K difference would take $5400/(4.18 \times 5) \approx 258\,\mathrm{kg/s}$, about a quarter of a cubic meter of water every second.

The water-side heat is *calculated*, from a flow meter and two temperature sensors, all *measured*. Its accuracy is limited by the temperature difference more than by either temperature. If each sensor is good to $\pm 0.2\,\mathrm{K}$, their difference can be off by up to 0.4 K, which is 8 % of a 5 K difference; a flow-meter error adds to that. A small temperature difference is a large relative uncertainty, and the water-side heat is correspondingly less certain than the electrical reference. Section 5.6 returns to the consequence: when the air side, the water side and the electrical meters disagree, the measurement chain itself is one of the suspects.

5.3.4 Why liquid cooling is attractive

Equations (5.5) and (5.10) differ only in the fluid. Per kilogram, water holds about four times as much heat per kelvin as air, $c_{p,w} \approx 4.18$ against $c_{p,\mathrm{air}} \approx 1.0\,\mathrm{kJ/(kg\,K)}$. Per unit of volume, which is what sets the size of pipes and ducts, the difference is far larger, because water is about 800 times denser: air at about $1.2\,\mathrm{kg/m^3}$ holds about $1.2\,\mathrm{kJ/(m^3\,K)}$, water at $1000\,\mathrm{kg/m^3}$ about $4180\,\mathrm{kJ/(m^3\,K)}$, roughly 3500 times as much.

For the 40 kW rack with a 10 K rise, air needs 4 kg/s, about $3.3\,\mathrm{m^3/s}$; water needs $40/(4.18 \times 10) \approx 0.96\,\mathrm{kg/s}$, under one liter per second. At a typical pipe velocity of 1 m/s that water fits a pipe of about 3.5 cm diameter, while the air at 2 m/s needs a flow area of about $1.7\,\mathrm{m^2}$. Large heat loads move through small pipes. The point matters most for high-density racks of graphics processing units (GPUs), the accelerators used for machine learning, whose rack powers of 100 kW and more (illustrative) would require 10 kg/s of air per rack at a 10 K rise: more air than the aisle, the servers' fans and the acoustics can practically deliver.

5.3.5 Direct-to-chip cooling

In an air-cooled server the heat of a processor follows the path chip → heat sink → air → coil → water. Direct-to-chip cooling replaces the first part of it: a cold plate, a metal block with internal channels, sits on the processor and liquid flows through it, so the path becomes chip → cold plate → liquid. The liquid then goes to a heat exchanger and into the facility water. The air stage, the stage with the weakest heat-carrying fluid, is removed from the path of the largest heat source.

The benefit is a lower thermal resistance, in the sense defined in Section 5.4.1: heat flow equals temperature difference over resistance, $\dot{Q} = \Delta T / R_{\mathrm{th}}$, so the same heat crosses a lower resistance with a smaller temperature difference. For an illustrative processor dissipating 700 W, an air path from chip to inlet air with $R_{\mathrm{th}} = 0.08\,\mathrm{K/W}$ needs a 56 K difference, while a cold plate with $R_{\mathrm{th}} = 0.02\,\mathrm{K/W}$ needs 14 K. With the chip held at the same maximum temperature, the coolant can be much warmer than the air would have to be, which lets the facility make cold fluid less often or not at all. Direct-to-chip cooling usually captures most but not all of a server's heat (for example 80 %, illustrative); memory, power supplies and drives still release the rest into air, so a liquid-cooled hall still has an air balance, smaller than before.

5.4 Resistance, capacitance and time

The balances so far relate flows to temperature differences and storage to temperature changes. This section names the two coefficients that do that, thermal resistance and thermal capacitance, shows that their product is a time, and uses that time to estimate how long a facility has when its cooling fails.

5.4.1 Thermal resistance and the electrical analogy

Heat flows from hot to cold at a rate proportional to the temperature difference along its path. The constant of proportionality is a conductance; its inverse is the thermal resistance,

$$\dot{Q} = \frac{T_1 - T_2}{R_{\text{th}}}, \tag{5.11}$$

in kelvin per watt. The law has exactly the form of Ohm's law for a resistor, $I = (V_1 - V_2)/R$: temperature plays the role of voltage, heat flow the role of current, and thermal resistance the role of electrical resistance. The analogy carries over the rules for combining resistances. Resistances in series add: the path from a chip to the air crosses the chip package, the thermal paste, the heat sink and the boundary layer of air, and its total resistance is the sum. Conductances in parallel add: two heat paths out of the same body remove heat in proportion to their conductances. The overall conductance UA of a heat exchanger used in Chapter 7 is one such conductance, $UA = 1/R_{\text{th}}$.

Storage completes the analogy. A thermal capacitance stores heat as a temperature, as an electrical capacitor stores charge as a voltage, and the balance of equation (5.2) is Kirchhoff's current law at a node: the heat arriving at a node either charges its capacitance or flows on through resistances. A body of capacitance C_{th} heated by $\dot{Q}$ and connected to a surrounding temperature T_{amb} through R_{th} obeys

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q} - \frac{T - T_{\text{amb}}}{R_{\text{th}}}, \qquad \text{against} \qquad C \frac{\mathrm{d}V}{\mathrm{d}t} = I - \frac{V}{R} \tag{5.12}$$

for a current source I feeding a capacitor C and a resistor R in parallel. The surrounding temperature plays the role of the electrical ground, the reference against which node voltages are measured.

Figure 5.4 draws both circuits. On the left is the electrical one: a current source (a circle with an arrow), a capacitor (two plates) and a resistor (a zigzag), each connected between a top node at voltage V and the ground rail. On the right is the thermal one, with the same topology: a heat source, the thermal capacitance and the thermal resistance between a node at temperature T and a rail at T_{amb}. The column in the middle lists the correspondence. The use of the analogy is practical: a facility with many temperatures becomes a network of such nodes, a reduced thermal-network model, and every tool for solving circuits, steady or transient, applies to it.

5.4.2 Thermal capacitance

The capacitance of a body is its mass times its specific heat,

$$C_{\text{th}} = m\, c_p. \tag{5.13}$$

An illustrative 1000 kg of equipment, mostly steel with $c_p \approx 0.5\,\text{kJ/(kg K)}$, has

$$C_{\text{th}} = 1000\,\text{kg} \times 0.5\,\text{kJ/(kg K)} = 500\,\text{kJ/K} :$$

adding 500 kJ of heat raises its temperature by about 1°C, and removing 500 kJ lowers it by the same amount.

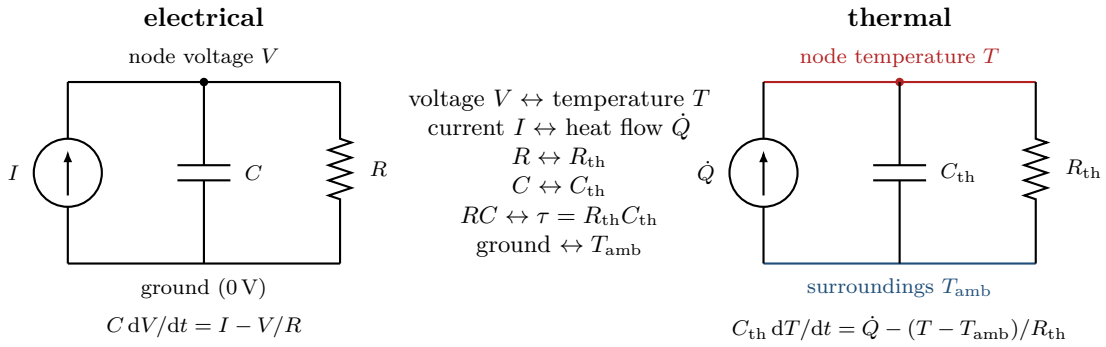


Figure 5.4: The electrical RC circuit (left) and the thermal circuit of one heated body (right), equation (5.12). The two have the same topology and the same equation; the middle column lists the correspondence. The surrounding temperature plays the role of ground.

A data center holds several kinds of capacitance with very different sizes. The air of an illustrative hall of 40 000 m^3, at about 1.2 kg/m^3, has a mass of 48 000 kg and a capacitance of only 48 MJ/K. The IT equipment, racks, cable trays and raised floor are hundreds of tonnes of metal; the concrete structure is thousands of tonnes; and the water in the chilled-water piping and any storage tanks holds 4.18 MJ/K per cubic meter. These masses are not equally well coupled to the air: the air exchanges heat with the equipment in seconds to minutes, with the building structure over hours. The *effective* capacitance that governs a room's temperature therefore depends on the time scale of the question, a point Section 5.4.4 returns to.

5.4.3 The thermal time constant

Equation (5.12) has one time scale. For a constant heat input $\dot{Q}$ and surrounding temperature T_{amb}, starting from T_0, its solution is

$$T(t) = T_{\mathrm{amb}} + \dot{Q}R_{\mathrm{th}} + (T_0 - T_{\mathrm{amb}} - \dot{Q}R_{\mathrm{th}})\, e^{-t/\tau}, \qquad \tau = R_{\mathrm{th}}C_{\mathrm{th}}. \tag{5.14}$$

The temperature approaches the equilibrium $T_{\mathrm{amb}} + \dot{Q}R_{\mathrm{th}}$ exponentially; after one time constant it has covered $1 - e^{-1} \approx 63\,\%$ of the distance, after three time constants 95 %. The units check: (K/W) × (J/K) = J/W = s. For the 500 kJ/K block of Section 5.4.2 coupled to the air through an illustrative $R_{\mathrm{th}} = 0.002$ K/W (a conductance of 500 W/K),

$$\tau = 0.002\,\mathrm{K/W} \times 500\,000\,\mathrm{J/K} = 1000\,\mathrm{s} \approx 17\,\mathrm{min}:$$

if the surrounding air steps up by 2 K, the block is 1.26 K warmer after 17 minutes and 1.9 K warmer after 50.

The time constants of a data center separate its physics into layers. Electrical disturbances, a fault, a transfer between sources, a step of load at a converter, play out in milliseconds; a cycle of the alternating supply lasts 16.7 ms at 60 Hz. Thermal changes take seconds at a processor die, minutes for the air of a room and the equipment in it, and hours for the building structure. An electrical event is over before any temperature responds, and a thermal event unfolds slowly enough to be observed, predicted and acted on. That difference of time scales is what makes the thermal side of a facility both forgiving and diagnosable.

5.4.4 Thermal inertia and the loss-of-cooling rise

Thermal inertia is valuable. If cooling disappears entirely, the removal term in equation (5.2) vanishes and all the IT heat is stored,

$$C\,\frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\mathrm{IT}}, \qquad \frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\dot{Q}_{\mathrm{IT}}}{C}. \tag{5.15}$$

The temperature rises at a rate inversely proportional to the capacitance. A larger capacitance gives a slower rise, and a slower rise gives operators and automatic controls time to respond: to start a standby chiller, to wait out a generator start, to shed load. Thermal mass is a short-term thermal buffer, and chilled-water storage tanks exist to enlarge it deliberately. The same equation applies to a partial loss, with $\dot{Q}_{\mathrm{IT}}$ replaced by the shortfall, the generation minus whatever removal remains.

Example 5.2. A facility with an illustrative effective capacitance $C_{\mathrm{eff}} = 500\,\mathrm{MJ/K}$ loses 2 MW of its cooling, for instance one of several chillers. By equation (5.15)

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{2\,\mathrm{MW}}{500\,\mathrm{MJ/K}} = 0.004\,\mathrm{K/s} = 0.24^{\circ}\mathrm{C}\ \text{per minute},$$

and after 10 minutes the temperature is 2.4°C higher. If the rack inlets start 5 K below their limit, they reach it after $5/0.24 = 20.8$ minutes. With half the capacitance the rise is twice as fast and the margin lasts 10.4 minutes.

Figure 5.5 plots the example. The horizontal axis is the time since the loss of cooling, in minutes; the vertical axis is the temperature rise in kelvin. The solid line is the 500 MJ/K facility, the dashed line a facility with 250 MJ/K, both under the same 2 MW shortfall. The dashed horizontal line is an illustrative 5 K margin to the inlet limit, and the vertical dotted lines mark the times at which each facility uses it up. Capacitance does not change how far the temperature will go if nothing is done; it changes how long there is to do something.

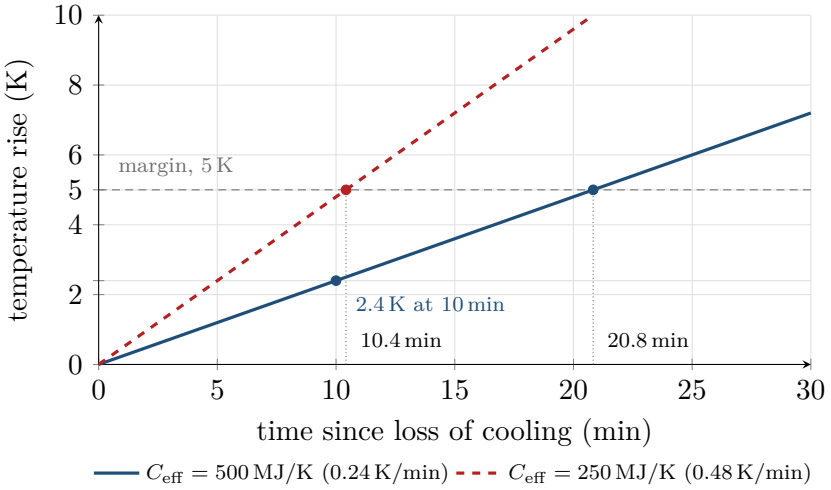


Figure 5.5: Temperature rise after a loss of 2 MW of cooling, $\Delta T(t) = \dot{Q}\,t/C_{\mathrm{eff}}$ (equation (5.15)), for two illustrative effective capacitances. Halving the capacitance halves the time to use up a 5 K margin, from 20.8 to 10.4 minutes.

The single-capacitance picture is a first estimate. Real facilities have multiple masses with different couplings and temperature gradients between them. In the first minutes after a loss of cooling only the air and the equipment it touches respond, so the effective capacitance is much closer to the 48 MJ/K of the air plus a part of the equipment than to the full mass of the building; the air near the racks warms first, and the rack inlets, which are what the limit applies to, warm fastest. A single number that counts slowly coupled masses therefore overstates the time available, and an honest estimate either uses the capacitance that couples on the time scale of interest or uses a network of several nodes, as in Section 5.5.1.

Principle 5.3. *Thermal capacitance converts a cooling shortfall into a rate of temperature rise,* $\mathrm{d}T/\mathrm{d}t = \text{shortfall}/C$. *It sets the time available to respond, and only the capacitance that is coupled to the air on that time scale counts.*

5.5 Spatial structure and mixing

A room is not one temperature. This section turns the single balance into a network of balances, describes the two mixing paths that make the air at a server inlet differ from the air a cooling unit supplies, shows why an average temperature hides them, and assembles a three-node model that can be run in steady state or in time.

5.5.1 The thermal network

A data center has many temperatures that differ by several kelvin at the same moment: rack inlet, rack outlet, cold aisle, hot aisle, ceiling return, CRAH inlet and CRAH outlet, besides the water temperatures of the loop. Each belongs to a different region of air, and each region can be taken as its own control volume with its own balance. For node i,

$$C_i \frac{\mathrm{d}T_i}{\mathrm{d}t} = \sum_j \dot{Q}_{j\to i} - \sum_k \dot{Q}_{i\to k} + \dot{Q}_{i,\mathrm{gen}}, \tag{5.16}$$

where $\dot{Q}_{j\to i}$ is the heat carried into node i from each neighbor j, $\dot{Q}_{i\to k}$ the heat carried out to each neighbor k, and $\dot{Q}_{i,\mathrm{gen}}$ any generation inside the node. The set of nodes and the heat flows between them form a *thermal network*.

The heat flows on the edges of the network are of two kinds. An *advective* edge carries heat with a moving fluid: air of mass flow $\dot{m}$ leaving a node at temperature T_i carries $\dot{m}c_pT_i$ (measured from a common reference), so the net heat a stream delivers between two nodes is $\dot{m}c_p(T_i - T_j)$, the form of equation (5.5). A *conductive* edge carries heat through a solid or a boundary layer, $(T_i - T_j)/R_{\mathrm{th}}$, the form of equation (5.11). Every model in the rest of the book, from one rack to a whole hall, is a network of this kind with more or fewer nodes.

5.5.2 Recirculation

The air a server draws in is not always the air the cooling unit supplied. Some hot exhaust may loop back to the rack inlet, over the top of a rack, around the end of a row, or through gaps between servers. Model the inlet air as a mixture of supply air at T_{supply} and hot exhaust at T_{hot}, with a recirculation fraction α equal to the share of inlet air that is exhaust. Mixing air streams of equal specific heat gives the mass-weighted mean temperature,

$$T_{\mathrm{in}} = (1-\alpha)\,T_{\mathrm{supply}} + \alpha\,T_{\mathrm{hot}}. \tag{5.17}$$

With $\alpha = 0$ there is no recirculation and the inlet receives supply air; with $\alpha = 0.2$, 20 % of the inlet air is hot exhaust.

Example 5.3. With $T_{\mathrm{supply}} = 20°\mathrm{C}$, $T_{\mathrm{hot}} = 40°\mathrm{C}$ and $\alpha = 0.2$,

$$T_{\mathrm{in}} = 0.8 \times 20 + 0.2 \times 40 = 16 + 8 = 24°\mathrm{C}.$$

The cooling unit supplies 20°C; the server sees 24°C.

Figure 5.6 draws the paths of air that produce recirculation and its counterpart, bypass. The tall box on the left is the cooling unit. Air leaves it at T_{supply} along the top row, passes a split point and a mixing point, goes through the rack, and reaches the hot aisle at T_{hot}; it returns along the bottom row through a second mixing point to the cooling unit at T_{return}. The red dashed path above the top row is recirculation: exhaust drawn back from the hot aisle into the rack inlet mix, a fraction α of the inlet flow. The blue dashed path is bypass: supply air that turns back at the split without reaching the rack, a fraction β of the supply. The two circles are the mixing points, each labeled with the temperature its mixture produces.

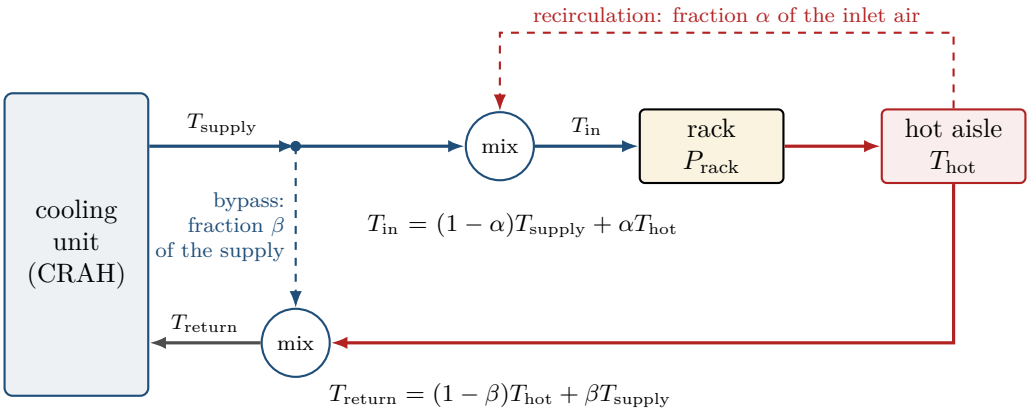


Figure 5.6: The two mixing paths of a room's air. Recirculation (red, dashed) returns hot exhaust to the rack inlet and raises T_{in} above T_{supply}, equation (5.17); bypass (blue, dashed) returns supply air to the cooling unit without passing through the rack and lowers T_{return} below T_{hot}, equation (5.18).

Figure 5.7 plots equation (5.17) against the recirculation fraction. The horizontal axis is α, the vertical axis the inlet temperature in degrees Celsius. The solid line is the equation with the exhaust temperature held at 40°C, as in Example 5.3; the dot marks $\alpha = 0.2$, 24°C. The dashed curve treats the exhaust temperature as the inlet temperature plus the rack's own rise, $T_{\text{hot}} = T_{\text{in}} + \Delta T$ with $\Delta T = 20\,\text{K}$, which is what happens when the recirculated air is the rack's own exhaust: a warmer inlet makes a warmer exhaust, which makes a warmer inlet. Solving equation (5.17) with that substitution gives $T_{\text{in}} = T_{\text{supply}} + \alpha\Delta T/(1-\alpha)$, which at $\alpha = 0.2$ is 25°C rather than 24°C. The gray dashed line is the upper end of a widely used recommended inlet range, about 27°C, here an illustrative limit. The fixed-exhaust line crosses it at $\alpha = 0.35$, the self-consistent curve at $\alpha = 7/27 \approx 0.26$: the feedback shortens the margin, and the inlet rise accelerates as recirculation grows.

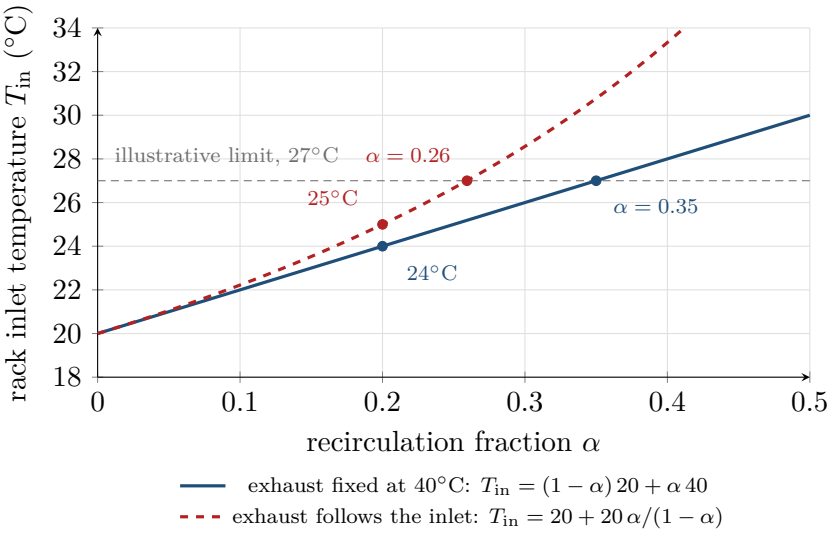


Figure 5.7: Rack inlet temperature against recirculation fraction, equation (5.17), for $T_{\text{supply}} = 20$°C (illustrative). Solid: exhaust held at $T_{\text{hot}} = 40$°C, giving 24°C at $\alpha = 0.2$. Dashed: exhaust equal to inlet plus a 20 K rack rise, $T_{\text{in}} = T_{\text{supply}} + \alpha\Delta T/(1-\alpha)$, giving 25°C at $\alpha = 0.2$ and reaching the 27°C illustrative limit at a smaller fraction.

Recirculation is not measured. Supply and inlet temperatures are measured; the hot-aisle

temperature is measured or calculated from the inlet and the rack's rise; and α is *inferred* by inverting equation (5.17), $\alpha = (T_{\text{in}} - T_{\text{supply}})/(T_{\text{hot}} - T_{\text{supply}})$, which is only as good as the three temperatures and the assumption that the inlet sensor sees a well-mixed sample.

5.5.3 Bypass airflow

Bypass is the opposite path: cold supply air that returns to the cooling unit without passing through any IT equipment, through perforated tiles in the wrong place, through open cable cutouts, or because the units supply more air than the racks draw. With a bypass fraction β of the supply, the air returning to the unit is a mixture of rack exhaust and supply air,

$$T_{\text{return}} = (1 - \beta)\, T_{\text{hot}} + \beta\, T_{\text{supply}}. \tag{5.18}$$

For an illustrative case with $T_{\text{supply}} = 20°\text{C}$, a rack exhaust of 35°C and $\beta = 0.3$, $T_{\text{return}} = 0.7 \times 35 + 0.3 \times 20 = 30.5°\text{C}$. The cooling unit sees a 10.5 K difference across its coil instead of the 15 K the racks produce.

Bypass uses cooling capacity inefficiently. The heat removed by the unit is its airflow times $c_{p,\text{air}}$ times its own temperature difference; with a lower effective difference it must move more air for the same heat, here $1/(1 - \beta) = 1.43$ times the air the racks actually use, and the fan power spent on the bypassed 30 % does no cooling. A smaller difference across the coil also lowers the coil's effectiveness and, in many plants, the chilled-water return temperature, which makes the chiller work harder. Bypass has a second consequence for diagnosis: a low return-air temperature looks like a room that is well cooled, but it can indicate bypass rather than good cooling. The same reading has opposite interpretations depending on a fraction that no sensor measures.

5.5.4 Why average temperature misleads

A low average room temperature does not show that the room is well cooled. An average over a room's sensors can hide overcooling in one area, where supply air is wasted, and hot spots elsewhere, where racks draw recirculated air; it can hide bypass, which pulls the average down, and recirculation, which affects only a few inlets; and it can hide uneven loading, a few dense racks among many light ones. Each of these moves the average by little and moves individual inlets by a lot. Ten racks at a 22°C inlet and one at 30°C average 22.7°C, a room that looks comfortable with one rack above an illustrative 27°C limit.

The question that matters to the IT equipment is narrower and local: are the rack inlet temperatures within limits? Server manufacturers specify inlet conditions, not room averages, and the inlet is where the equipment draws the air that cools it. Temperatures elsewhere are useful as evidence about the network, the supply, the return and the aisles, but the constraint is written on the inlets. A monitoring system that reports one room temperature, or one return temperature per cooling unit, answers a question that the equipment does not ask.

> **Principle 5.4.** *Limits apply to rack inlet temperatures, not to an average. A low average or a low return temperature is consistent with hot inlets, and with bypass rather than good cooling.*

5.5.5 A three-node thermal model

The smallest network that separates supply, equipment and exhaust has three nodes: the cold aisle at T_c, the rack at T_r and the hot aisle at T_h, with thermal capacitances C_c, C_r and C_h.

Writing equation (5.16) for each,

$$C_c \dot{T}_c = \dot{Q}_{\mathrm{CRAH}\to c} - \dot{Q}_{c\to r}, \tag{5.19}$$

$$C_r \dot{T}_r = P_{\mathrm{IT}} + \dot{Q}_{c\to r} - \dot{Q}_{r\to h}, \tag{5.20}$$

$$C_h \dot{T}_h = \dot{Q}_{r\to h} - \dot{Q}_{h\to \mathrm{CRAH}}, \tag{5.21}$$

where a dot over a temperature denotes its rate of change. The cooling unit delivers air into the cold aisle, the rack draws it, adds its electrical power as heat, and exhausts it into the hot aisle, from which it returns to the unit. With advective edges and rack airflow $\dot{m}_{\mathrm{air}}$, the edge flows are $\dot{Q}_{\mathrm{CRAH}\to c} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} T_{\mathrm{supply}}$, $\dot{Q}_{c\to r} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} T_c$, $\dot{Q}_{r\to h} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} T_r$ and $\dot{Q}_{h\to \mathrm{CRAH}} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} T_h$, measured from a common reference; at steady state the equations give $T_c = T_{\mathrm{supply}}$, $T_r = T_c + P_{\mathrm{IT}}/(\dot{m}_{\mathrm{air}} c_{p,\mathrm{air}})$, which is equation (5.6), and $T_h = T_r$. Recirculation and bypass add two edges with their own fractions, from the hot aisle to the cold aisle and from the cold aisle back to the unit.

Figure 5.8 draws the model. The three circles are the nodes, each labeled with its temperature and capacitance; the box on the left is the cooling unit, which closes the loop. Solid arrows are the edge flows of equations (5.19)–(5.21), the arrow from above is the IT power entering the rack node as generation, and the two dashed arrows are the recirculation and bypass edges. This is a reduced-order physical model: three temperatures instead of a full computational fluid dynamics solution of the room, but each equation is a conservation law and each parameter has a physical meaning, so the model can be checked against measurements and its parameters estimated from them.

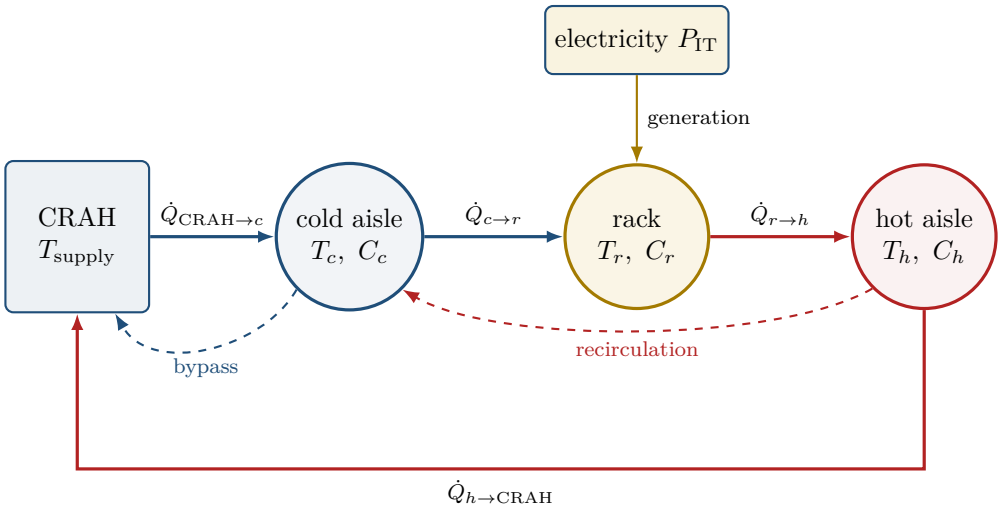


Figure 5.8: The three-node thermal model of equations (5.19)–(5.21): cold aisle, rack and hot aisle, each a conservation node with its own capacitance, closed through the cooling unit. Dashed edges are the recirculation and bypass paths that a fuller model adds.

5.5.6 Steady state and transients

At steady state every derivative in equations (5.19)–(5.21) is zero and each node satisfies $\dot{Q}_{\mathrm{in}} = \dot{Q}_{\mathrm{out}}$, counting generation as an inflow. Steady-state models answer sizing questions: how much cooling a room needs, what inlet temperature a given supply and recirculation produce, how much airflow a rack requires.

Many of the questions that matter operationally are transient. A sudden increase of GPU load as a training job starts, a fan failure in a rack, a CRAH shutdown, an interruption of chilled-water flow, the transition to generator power during which chillers stop and restart,

and a spike of outdoor temperature that reduces the plant's heat rejection: each changes one term of the balance faster than the temperatures can follow. The question a transient model answers is the one of Example 5.2: how much time remains before a temperature becomes unsafe. Steady-state analysis cannot answer it, because at steady state there is no time. Chapter 8 develops the transient models in detail.

5.6 Closing the balance and inference

A balance that should close and does not is a measurement. This section treats the gap between generated and removed heat as evidence, separates storage from sensor error with a time series, admits that the parameters of the model are uncertain, and writes the model as a probability over states, parameters and causes given the measurements.

5.6.1 An imbalance and its hypotheses

Take an illustrative facility with $P_{\mathrm{IT}} = 4.0\,\mathrm{MW}$ and 0.2 MW of other thermal load. By equation (5.8) the expected heat to remove is 4.2 MW. The chilled-water loop, from its flow meter and its supply and return temperatures by equation (5.10), shows 3.3 MW. The balance is short by 0.9 MW, more than a fifth of the load. Five hypotheses explain the gap, and Table 5.1 lists them with the size of the error each would need and the evidence that would separate it from the others.

Table 5.1: Five hypotheses for a 0.9 MW gap between the expected 4.2 MW and the 3.3 MW calculated from the chilled-water loop.

Hypothesis	What it would require	Evidence that separates it
H_1: flow-meter bias	the meter reads $3.3/4.2 = 79\,\%$ of the true flow, about 21 % low	a second flow measurement (clamp-on meter), pump curve and speed
H_2: temperature-sensor bias	at 200 kg/s the true difference is 5.0 K and the measured one 3.9 K: a 1.1 K error in the difference	swap or recalibrate sensors; compare with chiller-side temperatures
H_3: unmodeled heat path	0.9 MW leaves by another route: units with their own refrigeration, envelope losses, an unmetered liquid loop	an inventory of cooling paths; the air-side balance of each unit
H_4: IT power meter error	the IT meters read about 22 % high	cross-check against upstream switchgear and UPS meters
H_5: transient storage	the room and equipment are warming, absorbing 0.9 MW	temperature trends over time, Section 5.6.3

The table shows why balance closure is diagnostically valuable. Each hypothesis is a specific, quantitative claim: a flow meter 21 % low, a temperature difference 1.1 K low, 0.9 MW of missing heat path, meters 22 % high, or temperatures rising at a definite rate. Two of them are small-looking errors with large consequences: a 1.1 K error on a 5 K difference, as Section 5.3.3 noted, is easily produced by two sensors each within a modest tolerance. A balance that does not close points to a fault in the plant, in the instruments or in the model, and in each case it says how large the fault must be.

5.6.2 Storage as an explanation

Generation of 4.2 MW against removal of 3.3 MW is not a violation of conservation if the difference is being stored. By equation (5.2),

$$\dot{Q}_{\text{gen}} - \dot{Q}_{\text{removed}} = \frac{\mathrm{d}E_{\text{thermal}}}{\mathrm{d}t}, \tag{5.22}$$

and the 0.9 MW may be the rate at which the building and its equipment are warming. A balance that equates generation with removal is the steady-state balance; the full balance says that generation minus removal equals the storage rate, and an imbalance can be a transient rather than an error.

That relation must be respected in both directions. A model that attributes every imbalance to sensor bias will misread every warm-up and cool-down as a calibration problem. A model that attributes every imbalance to storage will miss a failing meter for as long as the temperatures happen to drift. Storage is one hypothesis among five, and it is the only one that makes a prediction about temperatures over time.

5.6.3 Time series separate storage from bias

A single snapshot cannot separate sensor bias from storage: at one instant, 4.2 MW generated and 3.3 MW removed are equally consistent with a 0.9 MW meter error and with 0.9 MW going into thermal mass. A temperature record $T(t)$ can. If the gap is storage, temperatures must be rising at the rate the gap implies; with the illustrative $C_{\text{eff}} = 500\,\text{MJ/K}$ of Example 5.2,

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{0.9\,\text{MW}}{500\,\text{MJ/K}} = 0.0018\,\text{K/s} \approx 0.11\,\text{K per minute} \approx 6.5\,\text{K per hour}.$$

If temperatures do rise at about that rate, the balance is correct and the gap is storage. If they hold within 0.5 K over an hour, storage can account for at most $0.5\,\text{K} \times 500\,\text{MJ/K}/3600\,\text{s} \approx 0.07\,\text{MW}$ of the gap, and the remaining 0.83 MW must be one of the other four hypotheses.

The argument depends on the capacitance and on which temperatures are used, since air warms faster than steel, and a careful version uses the network of Section 5.5.1 rather than one number. The principle does not depend on those details: dynamic data resolve ambiguities that static data cannot, because storage is the one explanation whose signature is a slope.

Principle 5.5. *At one instant a heat imbalance cannot be told apart from a measurement error. Over time it can: storage predicts a temperature slope of imbalance over capacitance, and a bias does not.*

5.6.4 Uncertain parameters

The parameters of the thermal model are not known precisely. The thermal resistance of a path, the effective capacitance of a room, the airflow through a rack and the recirculation fraction α at its inlet are all estimates: from design documents, from fan curves, from past data. Writing $R_{\text{th}} = 0.10\,\text{K/kW}$ states more than anyone knows. A more honest statement is a probability distribution, $R_{\text{th}} \sim p(R_{\text{th}})$, centered near 0.10 K/kW with a spread that reflects how well it is known.

Uncertain parameters make uncertain predictions. A forward simulation of the network with parameters drawn from their distributions produces not one future temperature but a distribution $p(T_{t+h})$ at each horizon h. For Example 5.2, if the effective capacitance is known only to lie between 400 and 600 MJ/K, the time to use up the 5 K margin lies between $5 \times 400/2 = 1000\,\text{s}$ and $5 \times 600/2 = 1500\,\text{s}$, between 16.7 and 25 minutes, instead of the single 20.8. An operating decision, such as whether a chiller can be restarted before the limit, then becomes a probability: the probability that the limit is reached before the restart completes. The deterministic balance has become a probabilistic model.

5.6.5 An inference formulation

The quantities of the model divide into three groups. Table 5.2 lists them with their status. The *state* $x = [T_{\text{cold}}, T_{\text{rack}}, T_{\text{hot}}, T_{\text{water}}, \dot{m}_{\text{air}}]$ holds the quantities that change in time and are governed by the balances; the *parameters* $\theta = [R_{\text{th}}, C_{\text{th}}, \alpha, \eta_{\text{coil}}]$ are the slowly varying coefficients, with η_{coil} the effectiveness of the coil, the ratio of the heat it transfers to the most it could transfer at the given inlet temperatures; and the *measurements* $y = [T_{\text{sensors}}, P_{\text{IT}}, \dot{m}_w, T_{\text{CHW,s}}, T_{\text{CHW,r}}]$ are what the instruments report.

Table 5.2: The inference formulation of the thermal model: states, parameters and measurements, with the status of each.

Group	Quantities	Status	Governed by
State x	T_{cold}, T_{rack}, T_{hot}, T_{water}, $\dot{m}_{\text{air}}$	inferred (some temperatures also measured)	node balances, equations (5.19)–(5.21) and (5.10)
Parameters θ	R_{th}, C_{th}, α, η_{coil}	inferred	prior knowledge $p(\theta)$, updated by data
Measurements y	temperature sensors, P_{IT}, $\dot{m}_w$, $T_{\text{CHW,s}}$, $T_{\text{CHW,r}}$	measured	sensor models with error and bias

The object of inference is the posterior distribution of the states and parameters given the measurements,

$$p(x, \theta \mid y) \;\propto\; p(y \mid x, \theta)\, p(x \mid \theta)\, p(\theta). \tag{5.23}$$

The first factor, the likelihood, says how probable the measurements are for a given state and parameters, and contains the sensor errors and biases; the second says which state trajectories the balances allow, and contains the physics; the third is the prior knowledge of the parameters. The posterior gives the estimated state together with its uncertainty: the most probable cold-aisle temperature and how far it may be off, the most probable recirculation fraction and its spread. A quantity that the measurements do not constrain keeps a posterior as wide as its prior, which is how the formulation reports a failure of identifiability such as the one of Section 5.2.4.

5.6.6 Probabilistic diagnosis of a rising inlet

The same formulation handles causes. Suppose a rack inlet temperature rises. Five hypotheses compete: H_1, the IT power increased (and more of a hotter exhaust recirculates); H_2, the rack airflow decreased; H_3, the recirculation increased; H_4, the supply temperature increased; H_5, the sensor is biased. The diagnosis is a probability for each,

$$\Pr(H_i \mid y) = \frac{p(y \mid H_i)\, \Pr(H_i)}{\sum_j p(y \mid H_j)\, \Pr(H_j)}, \tag{5.24}$$

where $\Pr(H_i)$ is how often each cause occurs and $p(y \mid H_i)$ how well it predicts all the measurements, not only the inlet.

The physics enters through $p(y \mid H_i)$. Each hypothesis predicts the same inlet rise but different effects on the other measurements, and Table 5.3 lists them: rack power, the rack's own temperature rise from inlet to outlet, the supply temperature, and the other sensors. A rise accompanied by higher rack power favors H_1; a larger rack temperature rise at unchanged power favors H_2; an inlet rise with unchanged supply, power and rack rise favors H_3; a supply rise that every rack served by the unit shares favors H_4; and a rise seen by one sensor alone, with

its neighbors and the rack outlet unchanged, favors H_5. The measurement that best separates the remaining hypotheses is the one to read, or install, next.

Table 5.3: Predicted signatures of five causes of a rising rack inlet temperature on the other measurements (↑ rises, = unchanged).

Hypothesis	Rack power	Rack rise $T_{\text{out}} - T_{\text{in}}$	Supply T_{supply}	Other sensors
H_1: IT power up	↑	↑	=	neighbors may rise slightly
H_2: rack airflow down	=	↑	=	fan speed or fan power changes
H_3: recirculation up	=	=	=	top-of-rack inlets rise more than bottom; return may fall
H_4: supply temperature up	=	=	↑	all racks served by the unit rise together
H_5: sensor biased	=	apparent fall	=	redundant sensor, neighbors and outlet unchanged

5.6.7 Equations to keep and the diagnostic question

Four relations carry most of the chapter:

$$\begin{aligned} \text{dynamic balance:} \quad & C\,\frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{generated}} - \dot{Q}_{\text{removed}}, \\ \text{steady balance:} \quad & \dot{Q}_{\text{generated}} = \dot{Q}_{\text{removed}}, \\ \text{air side:} \quad & \dot{Q} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, \Delta T, \\ \text{water side:} \quad & \dot{Q} = \dot{m}_{w}\, c_{p,w}\, \Delta T_{w}. \end{aligned}$$

With $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$, $\dot{Q} = \Delta T / R_{\text{th}}$ and $\tau = R_{\text{th}} C_{\text{th}}$ they reproduce every number of the chapter.

They also reframe the operational question. A data center thermal system is generation, transport and storage: servers generate heat, air and water transport it, thermal masses store it. Faced with a high temperature, the question is not why that temperature is high, but which of four things is true: generation is too high, transport is too weak, storage is changing, or the measurement is wrong. Each has its own balance term and its own signature in the other measurements, and the balance equations are what turn a reading into an answer.

5.7 What to retain

Definition 5.1 (Data center heat balance)**.** A data center thermal system is generation, transport and storage: servers turn electrical power into heat, $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$; air and water carry it, $\dot{Q} = \dot{m} c_p \Delta T$; and thermal masses store it, $C\,\mathrm{d}T/\mathrm{d}t = \dot{Q}_{\text{generated}} - \dot{Q}_{\text{removed}}$. A high temperature is explained by asking whether generation is too high, transport too weak, storage changing, or the measurement wrong.

Table 5.4 lists the chapter's objects and the subsection that defines each.

Table 5.4: The objects of the thermal balance and where each is defined.

Object	Expression	Role	Defined in
Energy balance	$C\,\mathrm{d}T/\mathrm{d}t = \dot{Q}_{\mathrm{in}} - \dot{Q}_{\mathrm{out}} + \dot{Q}_{\mathrm{gen}}$	conservation for any control volume	5.1.1
IT heat	$\dot{Q}_{\mathrm{IT}} \approx P_{\mathrm{IT}}$	generation, read from electrical meters	5.1.2
Air-side heat	$\dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} (T_{\mathrm{out}} - T_{\mathrm{in}})$	transport by air	5.2.2
Air temperature rise	$\Delta T = \dot{Q}_{\mathrm{rack}} / (\dot{m}_{\mathrm{air}} c_{p,\mathrm{air}})$	fixed by power and airflow	5.2.3
Room balance	equation (5.7)	cooling load from electrical terms	5.3.1
Water-side heat	$\dot{m}_w c_{p,w} (T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}})$	transport by chilled water	5.3.3
Thermal resistance	$\dot{Q} = (T_1 - T_2)/R_{\mathrm{th}}$	conduction; the electrical analogy	5.4.1
Thermal capacitance	$C_{\mathrm{th}} = m c_p$	storage	5.4.2
Time constant	$\tau = R_{\mathrm{th}} C_{\mathrm{th}}$	speed of thermal response	5.4.3
Loss-of-cooling rise	$\mathrm{d}T/\mathrm{d}t = \mathrm{shortfall}/C$	time available to respond	5.4.4
Recirculation fraction	$T_{\mathrm{in}} = (1-\alpha) T_{\mathrm{supply}} + \alpha T_{\mathrm{hot}}$	hot air at the inlet	5.5.2
Bypass fraction	$T_{\mathrm{return}} = (1-\beta) T_{\mathrm{hot}} + \beta T_{\mathrm{supply}}$	cold air at the return	5.5.3
Three-node model	equations (5.19)–(5.21)	reduced-order network	5.5.5
Imbalance hypotheses	Table 5.1	explanations of a balance gap	5.6.1
Posterior	$p(x, \theta \mid y)$	states and parameters with uncertainty	5.6.5
Diagnosis	$\Pr(H_i \mid y)$, Table 5.3	ranked causes of a rising inlet	5.6.6

5.8 Exercises

Exercise 5.1. A rack draws $P_{\text{rack}} = 32\,\text{kW}$ with an airflow of $2.5\,\text{kg/s}$ and an inlet at 22°C; take $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$. (a) Compute the air temperature rise and the outlet temperature. (b) Compute the outlet temperature if the airflow falls to $2.0\,\text{kg/s}$ at the same power, and if instead the power rises to $40\,\text{kW}$ at the same airflow. Which measurement would distinguish the two cases? (c) The rack is converted to liquid cooling with a water temperature rise of $8\,\text{K}$; take $c_{p,w} = 4.18\,\text{kJ/(kg K)}$. What water flow carries the $40\,\text{kW}$?

Exercise 5.2. A room has $P_{\text{IT}} = 3.2\,\text{MW}$, $0.2\,\text{MW}$ of electrical losses and $0.08\,\text{MW}$ of other gains. Its chilled-water loop carries $\dot{m}_w = 160\,\text{kg/s}$ with a measured difference $T_{\text{CHW,r}} - T_{\text{CHW,s}} = 5.0\,\text{K}$. (a) Compute the expected heat and the water-side heat, and the imbalance. (b) With an effective capacitance of $400\,\text{MJ/K}$, how fast would temperatures rise if the imbalance were storage? (c) Temperatures held within $0.2\,\text{K}$ over two hours. What is the most storage can explain? (d) If the rest is a flow-meter bias, by what percentage does the meter read low; if it is a temperature-difference bias, how large is it?

Exercise 5.3. A hall with an illustrative effective capacitance of $300\,\text{MJ/K}$ carries $3.6\,\text{MW}$ of load, and its rack inlets are $6\,\text{K}$ below their limit. (a) All cooling is lost. At what rate do temperatures rise, and when is the limit reached? (b) Chilled-water storage keeps removing $2.4\,\text{MW}$ during the outage. Repeat (a). (c) A chiller restart takes 10 minutes. In case (a), if the capacitance is known only to lie between 240 and $360\,\text{MJ/K}$, is the limit reached before the restart, and in case (b)?

Exercise 5.4. A rack inlet sensor rises from 22 to 26°C. The rack power is unchanged at $30\,\text{kW}$, the CRAH supply temperature is unchanged at 20°C, the rack's own temperature rise from inlet to outlet is unchanged at $12\,\text{K}$, a redundant inlet sensor agrees, and the inlet of the neighboring rack has also risen. (a) Use Table 5.3 to say which hypotheses survive. (b) Taking the exhaust temperature as the inlet plus $12\,\text{K}$, estimate the recirculation fraction before and after with equation (5.17). (c) Had the rack rise instead grown from 12 to $18\,\text{K}$ at unchanged power, which hypothesis would lead, and by what fraction would the airflow have changed?

5.9 Connection to the TERRA framework

The thermal balance is the third conservation network of a data center, beside electrical power and the mass flows of air and water, and it maps onto a graph node by node.

- **Control volumes are conservation nodes.** The rack, the cold aisle, the hot aisle, the room and the water loop are nodes whose stored quantity is $C_i T_i$; equation (5.16) is the balance at each, and equations (5.19)–(5.21) are the three-node instance of Figure 5.8.
- **Edges carry heat with mass or through resistance.** Advective edges carry $\dot{m} c_p \Delta T$ and couple the thermal network to the air and water mass networks; conductive edges carry $\Delta T / R_{\text{th}}$. Recirculation and bypass are edges whose fractions α and β are uncertain parameters, and the coil is a component that joins an air node to a water node with $\dot{Q}_{\text{air}} \approx \dot{Q}_{\text{water}}$.
- **The electrical network supplies the generation.** P_{IT}, measured on the power network, enters each rack node as generation by $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$, and compressor work enters at the chiller, so the electrical and thermal networks share variables rather than duplicating them.
- **Balances that should close are residuals.** The electrical, air-side and water-side estimates of the same heat are redundant relations; the $0.9\,\text{MW}$ gap of Table 5.1 is a

residual, and its five hypotheses are a bias on a flow edge, a bias on a temperature measurement, a missing edge, a bias on a power measurement, and a nonzero storage term. A time series separates them because only storage predicts a slope.

- **Inference returns a posterior, and a risk is a probability.** The posterior $p(x, \theta \mid y)$ of equation (5.23) estimates the unmeasured states and the inferred parameters R_{th}, C_{th}, α, η_{coil} and rack airflow; forward simulation with that posterior turns the time to a limit into a probability of reaching it before a restart.
- **Diagnosis is a ranked posterior over hypotheses.** $\Pr(H_i \mid y)$ for a rising inlet ranks the five causes of Table 5.3, and the next measurement is the one whose predicted values differ most across the hypotheses still in contention.

In summary, a data center's heat is generated by its electrical load, transported by air and water through a chain of balances that each should close, and stored in thermal masses that set the time available when the chain is interrupted; a temperature is a state of that network, and explaining it means deciding, with a posterior over states, parameters and hypotheses, whether generation, transport, storage or measurement has changed. The next chapter follows the air itself: Chapter 6 treats how air moves through the room, what drives it, and how containment controls the recirculation and bypass paths introduced here.

Chapter 6

Airflow and rack cooling: delivery, mixing, pressure and fans

6.1 Airflow required by a rack

Chapter 5 showed that the electrical power drawn by the IT equipment leaves it as heat, and that the facility exists to carry that heat out of the building. Air is the first medium that carries it: the servers reject their heat into air drawn through them, and that air must be supplied cool and taken away hot. This chapter studies that first transport step. Its question is

why one rack can overheat while the room as a whole is adequately cooled, and what the measurements of a data hall can determine about the air that reaches each rack.

The answer is that cooling is a matter of transport as much as of capacity: the right amount of cool air has to reach the right place, and the hot air has to be removed without mixing back into the supply. The chapter starts from the energy balance of one rack, follows the air through the aisles, the floor and the fans, and ends with the reduced network model, its relation to detailed fluid simulation, and the control and measurement decisions that the model supports.

Jargon: cold aisle, hot aisle, plenum, perforated tile, recirculation, bypass, containment, fan curve, system curve, affinity laws, CFD

Racks are placed in rows so that the air intakes (fronts) of two facing rows share a *cold aisle*, the corridor that receives cool supply air, and the exhausts (rears) of two back-to-back rows share a *hot aisle*, the corridor that collects the heated air. A *plenum* is an enclosed air volume used to distribute or collect air: the space under a raised floor through which supply air is pushed, or the space above a suspended ceiling through which hot air returns. A *perforated tile* is a floor panel with openings that lets supply air rise from an underfloor plenum into a cold aisle. *Recirculation* is hot exhaust air re-entering a rack inlet; *bypass* is cool supply air returning to the cooling unit without passing through any IT equipment. *Containment* is a physical enclosure (doors, roof panels, curtains) around a cold aisle or a hot aisle that keeps supply and exhaust air apart. A *fan curve* is the pressure rise a fan produces as a function of the volumetric flow it delivers at a given speed; a *system curve* is the pressure drop that the ducts, filters, coils, floor and racks impose as a function of the same flow. The *affinity laws* state how a fan's flow, pressure and power scale with its rotational speed. *Computational fluid dynamics* (CFD) is the numerical solution of the equations of fluid motion and heat transport on a fine spatial grid.

Key idea

A rack is cooled by the air that actually reaches its inlet, at the temperature that air actually has. Total cooling capacity in the room is necessary but not sufficient: the required airflow has to be delivered to every rack, through a network of pressures and resistances, without mixing with the exhaust. Cooling capacity is not cooling delivery.

The rest of this section writes the energy balance of one rack, works a numerical example, and separates global from local adequacy.

6.1.1 The rack airflow balance

Treat a rack as a control volume through which air flows from front to back. Air enters at temperature T_{in} with mass flow rate $\dot{m}_{\text{air}}$, absorbs the heat the equipment rejects, and leaves at temperature T_{out}. In steady state the energy carried out by the air equals the heat generated inside,

$$\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, (T_{\text{out}} - T_{\text{in}}), \tag{6.1}$$

where $c_{p,\text{air}} \approx 1.0\,\text{kJ/(kg K)}$ is the specific heat of air at constant pressure. The relation is the same sensible-heat balance used in Chapter 5; here it is read as a requirement on the air. Solving for the flow gives the airflow a rack needs to carry its heat with a given temperature rise $\Delta T = T_{\text{out}} - T_{\text{in}}$,

$$\dot{m}_{\text{air}} = \frac{\dot{Q}_{\text{rack}}}{c_{p,\text{air}}\, \Delta T}. \tag{6.2}$$

The required flow grows in proportion to the heat and in inverse proportion to the allowed temperature rise. A rack designed for a small rise needs a lot of air; a rack that tolerates a large rise needs less, at the cost of a hot exhaust.

Equation (6.1) assumes that all the electrical power becomes sensible heat in the air stream (no other heat path, no condensation or evaporation), and that T_{in} and T_{out} are the bulk temperatures of the entering and leaving streams. Section 6.5.3 returns to these assumptions when the balance is used in reverse, to infer the flow from temperatures.

6.1.2 A rack of 30 kW

Take an illustrative rack that draws $P_{\text{rack}} = 30\,\text{kW}$ and is designed for a temperature rise of 12 K. With $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$, equation (6.2) gives

$$\dot{m}_{\text{air}} = \frac{30\,\text{kW}}{1.0\,\text{kJ/(kg K)} \times 12\,\text{K}} = 2.5\,\text{kg/s}.$$

At an air density of about $1.2\,\text{kg/m}^3$ this is $2.5/1.2 \approx 2.08\,\text{m}^3\text{/s}$, or about $7500\,\text{m}^3\text{/h}$ (about 4400 cubic feet per minute, the unit common in North American practice). A rack is a large air consumer: the flow through it renews the volume of a small room every few seconds.

Suppose the floor in front of this rack delivers only 1.8 kg/s of supply air. The heat does not change, so something else must. The air leaving the aisle around the rack now has to carry 30 kW with 1.8 kg/s, so its net temperature rise over the supply is $30/(1.8 \times 1.0) = 16.7\,\text{K}$ instead of 12 K. How that larger rise appears depends on what the servers do. If the server fans keep pulling 2.5 kg/s, the missing 0.7 kg/s is drawn from wherever air is available, which is the hot exhaust: the inlet air becomes a mixture, and with a supply at 20°C the inlet settles near 24.7°C and the exhaust near 36.7°C (the calculation is in Section 6.2.3). If instead the flow through the servers falls to 1.8 kg/s, the inlet stays at the supply temperature and the exhaust rises by the full 16.7 K. In both cases the component temperatures inside the servers rise, because they sit downstream of a warmer inlet or in a slower stream, and in practice the server

fans respond to those temperatures by speeding up. The four symptoms of an airflow shortfall are therefore a higher exhaust temperature, a higher inlet temperature, higher component temperatures, and faster server fans, in any combination.

6.1.3 Global adequacy and local adequacy

Consider an illustrative hall with $\dot{Q}_{\text{avail}} = 5\,\text{MW}$ of installed cooling capacity and an IT load of 4 MW, so $\dot{Q}_{\text{req}} = 4\,\text{MW}$. Globally the room is adequately cooled, with a margin of 1 MW, or 25 % of the load. That statement is a comparison of two sums. It says nothing about any one rack. Rack 27, say, may sit at the far end of a row, behind an obstruction under the floor, or next to a gap through which exhaust returns, and receive too little cool air for its own heat. Writing the heat that the air delivered to rack i can absorb within its allowed temperature rise as $\dot{Q}_{\text{avail},i}$, the room can satisfy

$$\sum_i \dot{Q}_{\text{avail},i} \;\geq\; \sum_i \dot{Q}_{\text{rack},i} \qquad \text{while} \qquad \dot{Q}_{\text{avail},27} \;<\; \dot{Q}_{\text{rack},27}. \tag{6.3}$$

With the numbers of the previous subsection, if rack 27 is the 30 kW rack and receives 1.8 kg/s of supply air with a 12 K allowed rise, $\dot{Q}_{\text{avail},27} = 1.8 \times 1.0 \times 12 = 21.6\,\text{kW}$, short of its 30 kW by 8.4 kW, inside a room with 1 MW to spare.

The surplus elsewhere does not help rack 27, because heat does not move between racks by itself; it moves only where the air carries it. This is the same distinction as between total generation and the flow on one line of a transmission network: a system can have enough supply in aggregate and still fail locally because the network cannot deliver it where it is needed. Local adequacy, one inequality per rack, is the condition that matters, and the rest of the chapter is about the transport network that decides it.

> **Principle 6.1.** *Global cooling adequacy, $\dot{Q}_{\text{avail}} \geq \dot{Q}_{\text{req}}$ for the room, does not imply local adequacy, $\dot{Q}_{\text{avail},i} \geq \dot{Q}_{\text{rack},i}$ for every rack i. A rack is cooled only by the air that reaches it.*

6.2 Separation and mixing

Cool air is useful only while it is cool. This section describes the aisle arrangement that keeps supply and exhaust apart, quantifies the two ways the separation fails, recirculation and bypass, and shows why the average temperature of a room hides both.

6.2.1 The cold-aisle and hot-aisle arrangement

Servers draw air in at the front and blow it out at the back. Racks are therefore arranged in rows with alternating orientation: two rows face each other across a cold aisle, so both draw from the same supply, and two rows stand back to back across a hot aisle, so both exhaust into the same corridor. The intended path of the air is

$$\text{cool supply} \to \text{cold aisle} \to \text{rack inlet} \to \text{server} \to \text{hot aisle} \to \text{return} \to \text{cooling unit},$$

a loop in which the supply never meets the exhaust before the exhaust has been cooled again. The arrangement does not cool anything by itself; it organizes the paths so that every kilogram of supply air passes through a server once and every kilogram of exhaust goes back to the cooling unit once.

Figure 6.1 is a plan view, seen from above, of one such arrangement. The gray rectangles are racks in two rows; the thick blue edge of each rack is its front. The shaded blue band between

the rows is the cold aisle, with perforated floor tiles (dotted squares) where supply air rises from under the floor; the shaded red bands outside the rows are the two hot aisles. The cooling unit, a computer room air handler (CRAH) that cools air over a chilled-water coil, stands at the left. Solid arrows trace the intended path: supply from the CRAH into the cold aisle, through two of the racks, into the hot aisles and back to the CRAH. The two dashed paths at the right end are the two failures this section names: a red dashed arrow carries hot exhaust around the end of the upper row into the cold aisle (recirculation), and a blue dashed arrow carries cold supply around the end of the lower row into a hot aisle without passing through any server (bypass). Recirculation also happens over the tops of racks and through empty rack positions, which a plan view cannot show.

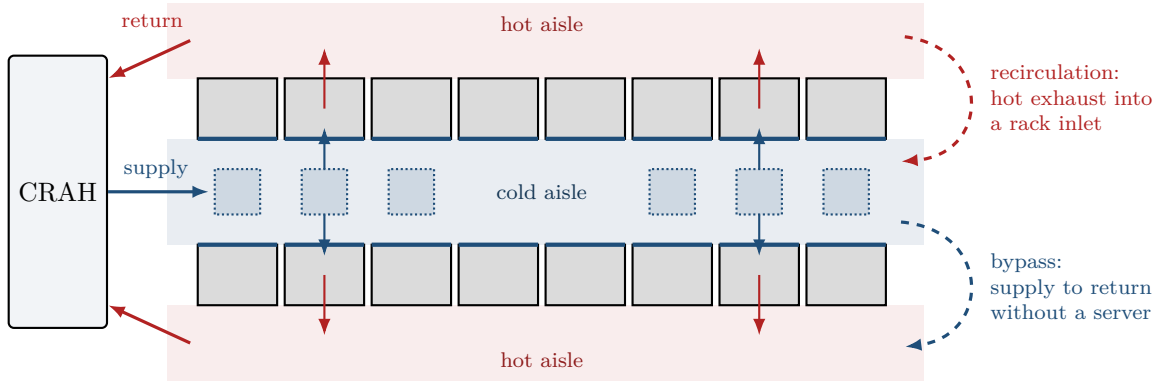


Figure 6.1: Plan view of a cold-aisle and hot-aisle arrangement, schematic. Two rows of racks (gray, fronts drawn as thick blue edges) face a shared cold aisle supplied through perforated tiles; their exhausts enter two hot aisles that return to the computer room air handler (CRAH). Solid arrows: the intended path. Dashed red: recirculation of exhaust around the end of a row into the cold aisle. Dashed blue: bypass of supply air around the end of a row into a hot aisle.

6.2.2 Mixing at the rack inlet

The reason to keep the streams apart is that a rack inlet takes whatever air is in front of it. Suppose the supply air leaves the cooling unit at $T_{\text{supply}} = 20\,^{\circ}\text{C}$ and the exhaust in the hot aisle is at $T_{\text{hot}} = 38\,^{\circ}\text{C}$. If 25 % of the air entering a rack is recirculated exhaust and 75 % is supply, the inlet air is a mixture, and its temperature is the mass-weighted average of the two streams (both have the same specific heat):

$$T_{\text{in}} = 0.75 \times 20\,^{\circ}\text{C} + 0.25 \times 38\,^{\circ}\text{C} = 15.0 + 9.5 = 24.5\,^{\circ}\text{C}.$$

The rack sees 24.5°C although the cooling unit is supplying 20°C air exactly as designed.

The example locates the problem. Nothing is wrong with the cooling unit, the chilled water or the chiller: the supply temperature is on its setpoint, and the plant delivers the capacity it was asked for. The 4.5 K excess at the inlet is produced entirely by the paths the air takes between the cooling unit and the rack, that is, by the airflow topology. A diagnosis that looks only at the cooling plant finds nothing, and a remedy that lowers the supply temperature treats the symptom at the cost of more chiller energy while the mixing continues.

6.2.3 The recirculation fraction

The example generalizes to a parameter. Let the air entering a rack be the sum of a supply stream $\dot{m}_{\text{sup}}$ and a recirculated stream $\dot{m}_{\text{rec}}$, and define the recirculation fraction as the share

of the inlet air that is exhaust,

$$\alpha = \frac{\dot{m}_{\text{rec}}}{\dot{m}_{\text{sup}} + \dot{m}_{\text{rec}}}, \qquad 0 \le \alpha \le 1. \tag{6.4}$$

Mass conservation and energy conservation at the inlet, with perfect mixing, give

$$T_{\text{in}} = (1 - \alpha)\, T_{\text{supply}} + \alpha\, T_{\text{hot}}. \tag{6.5}$$

The value $\alpha = 0$ is perfect separation, with the inlet at the supply temperature; $\alpha = 0.3$ means that 30 % of the inlet air is recirculated exhaust, and with the temperatures above it gives $T_{\text{in}} = 0.7 \times 20 + 0.3 \times 38 = 25.4°\text{C}$. Solved for the fraction, equation (6.5) reads

$$\alpha = \frac{T_{\text{in}} - T_{\text{supply}}}{T_{\text{hot}} - T_{\text{supply}}}, \tag{6.6}$$

so three temperatures give the recirculation fraction; for the example, $(24.5-20)/(38-20) = 0.25$. That number is calculated from measurements, but what it represents is inferred: the mixing is never observed directly, and the result depends on which hot-aisle temperature is taken to be the one that recirculates. In a reduced model of a data hall, α is a parameter per rack that compresses the whole three-dimensional flow around the rack into one number.

When the recirculated air is the rack's own exhaust, T_{hot} is not fixed: it is $T_{\text{in}} + \Delta T$, with ΔT the rise across the rack. Substituting into equation (6.5) and solving gives

$$T_{\text{in}} = T_{\text{supply}} + \frac{\alpha}{1 - \alpha}\, \Delta T, \tag{6.7}$$

which grows faster than linearly in α, because the recirculated air is itself warmed by recirculation. For the 30 kW rack of Section 6.1.2, which keeps pulling 2.5 kg/s while only 1.8 kg/s of supply arrives, $\alpha = 0.7/2.5 = 0.28$ and $T_{\text{in}} = 20 + (0.28/0.72) \times 12 = 24.7°\text{C}$, with exhaust at 36.7°C: the numbers quoted in that subsection.

Figure 6.2 plots both relations. The horizontal axis is the recirculation fraction α; the vertical axis is the rack inlet temperature. The solid line is equation (6.5) with a fixed hot-aisle temperature of 38°C, a straight line from the supply temperature at $\alpha = 0$ to the hot-aisle temperature at $\alpha = 1$; the dot marks the $\alpha = 0.25$, 24.5°C example. The dashed curve is equation (6.7) with $\Delta T = 12\,\text{K}$, for a rack that recirculates its own exhaust. The dotted red line is an illustrative inlet limit of 27°C, the upper end of the inlet range commonly recommended for air-cooled IT equipment. The solid line reaches the limit at $\alpha = 7/18 \approx 0.39$; the dashed curve, which bends upward, reaches it at $\alpha = 7/19 \approx 0.37$ and climbs steeply beyond. A moderate recirculation fraction is tolerable; a large one is not, whatever the supply temperature.

Principle 6.2. *Recirculation raises a rack's inlet temperature by $\alpha(T_{\text{hot}} - T_{\text{supply}})$ with no change in the cooling unit or the supply temperature. An inlet that is too hot is an airflow-topology problem before it is a cooling-plant problem.*

6.2.4 Bypass

The opposite failure is cool supply air that returns to the cooling unit without passing through any server: through perforated tiles placed in a hot aisle, through cable openings in the floor, around the ends of rows, or out of a cold aisle that receives more air than its racks draw. Define the bypass fraction as the share of the supplied air that does no cooling,

$$\beta = \frac{\dot{m}_{\text{byp}}}{\dot{m}_{\text{sup}}}. \tag{6.8}$$

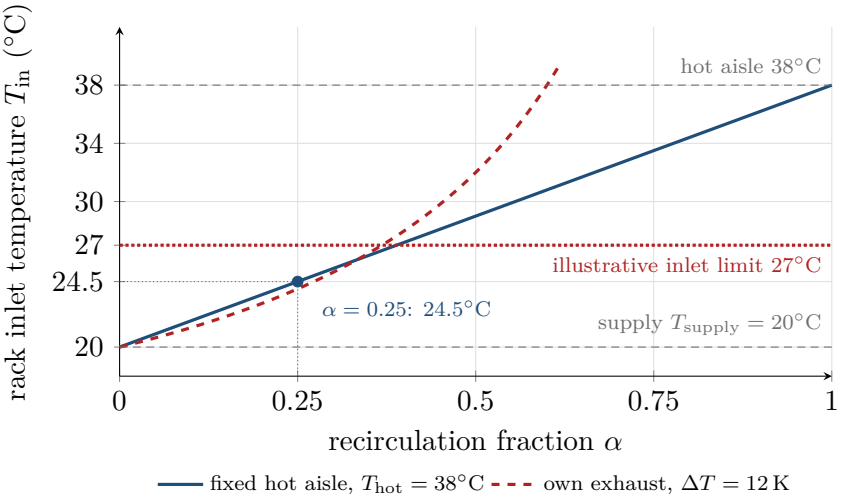


Figure 6.2: Rack inlet temperature against recirculation fraction, illustrative. Solid: $T_{in} = (1-\alpha)T_{supply} + \alpha T_{hot}$ (equation (6.5)) with $T_{supply} = 20$°C and $T_{hot} = 38$°C; the dot is the 25 % example. Dashed: $T_{in} = T_{supply} + \alpha\,\Delta T/(1-\alpha)$ (equation (6.7)) with $\Delta T = 12$ K, for a rack that recirculates its own exhaust. The supply temperature is the same everywhere on both curves.

If a CRAH supplies 100 kg/s and only 75 kg/s passes through the IT equipment, the other 25 kg/s is bypass and $\beta = 25/100 = 0.25$.

Bypass wastes fan energy, because a quarter of the air the fans move does no useful work, and it lowers the return temperature. With the IT equipment warming its 75 kg/s by 12 K, the return is a mixture of 75 kg/s at $T_{supply} + 12$ K and 25 kg/s at T_{supply}, which arrives at $T_{supply} + 0.75 \times 12 = T_{supply} + 9$ K. The cooling coil then works across a 9 K difference instead of 12 K. Its heat removal is the same, $75 \times 1.0 \times 12 = 900$ kW $= 100 \times 1.0 \times 9$, but a coil and a chiller generally transfer heat more effectively, and the plant can use outdoor air more often, when the return air is warmer; Chapter 7 develops that dependence.

6.2.5 Recirculation and bypass together

The two failures look opposite. Recirculation sends hot exhaust to a rack inlet, so an inlet is too hot. Bypass sends cold supply to the return, so airflow is wasted. A naive reading suggests that a hall has one or the other: too little air (recirculation) or too much (bypass). In practice both occur at the same time in the same hall, because the distribution of air is uneven. Tiles close to the cooling unit, or a cold aisle with too many open tiles for its racks, are oversupplied, and their excess spills to the return; racks at the far end of a row, or behind an obstruction under the floor, are starved, and their fans make up the shortfall from the hot aisle.

An illustrative hall shows both at once. The CRAH units supply 100 kg/s; 25 kg/s of it bypasses, so $\beta = 0.25$ and 75 kg/s reaches rack inlets. The racks' fans draw 90 kg/s in total. The 15 kg/s difference can come only from the hot aisles, so the hall-wide recirculation fraction is $15/90 \approx 0.17$. Adding supply air does not remove that recirculation efficiently: if a quarter of every additional kilogram also bypasses, the cooling units must deliver 20 kg/s more to put 15 kg/s more into the racks. The remedy for coexisting bypass and recirculation is to correct the distribution, not to increase the total.

6.2.6 The average room temperature

Because supply, exhaust and mixtures coexist in one room, a single room temperature says little. An illustrative hall reports an average temperature of 22°C from ten inlet sensors: three in well-supplied cold aisles read 18°C, five read 22°C, and two at the ends of starved rows read 28°C. The average is $(3 \times 18 + 5 \times 22 + 2 \times 28)/10 = 22$°C, comfortably within any limit, while two racks are above the illustrative 27°C limit and three are overcooled.

The constraint that protects equipment is written per rack, $T_{\text{in},i} \le T_{\max}$ for every rack i, and it is violated by the warmest inlet, not by the average. The room average $T_{\text{room,avg}}$ is therefore much less useful than the set of rack inlet temperatures $T_{\text{in},i}$, one per rack, and an alarm or a controller that acts on the average can report a healthy room while individual racks overheat. The lesson is the local-adequacy principle of Section 6.1.3 applied to temperature measurements.

6.3 Pressure, resistance and fans

Air moves only when a pressure difference pushes it through the resistances in its path. This section describes how air is distributed, the quadratic law that links pressure and flow, the fan and system curves whose intersection sets the flow, the scaling of fan power with speed, and the coupling between facility fans and server fans.

6.3.1 Air distribution paths

In a raised-floor hall the cooling units blow supply air down into the underfloor plenum, the space between the structural slab and the raised floor, which they pressurize slightly above the room. The air spreads through the plenum and rises through perforated tiles placed in the cold aisles, where the racks draw it in. The path is

$$\text{CRAH} \rightarrow \text{underfloor plenum} \rightarrow \text{perforated tile} \rightarrow \text{rack inlet.}$$

The return path runs the other way above head height: exhaust rises from the hot aisle, through grilles into a ceiling plenum or directly across the room, to the top of the cooling units. Other halls use overhead distribution, with supply ducts or diffusers above the cold aisles and no raised floor, or place the cooling units in the rows between racks. In every design the objective is the same: deliver the required airflow, equation (6.2), to every rack inlet.

Figure 6.3 is a section through a raised-floor hall, cut vertically across one cold aisle and one hot aisle. The gray band at the bottom is the slab; the line above it is the raised floor, with a dashed segment where perforated tiles let air through; the space between them is the underfloor plenum. The CRAH stands at the left, with its filter, coil and fan. Blue arrows show supply air pushed down into the plenum, along it and up through the tiles into the cold aisle and the rack front. The red arrow through the rack is the heated stream, which enters the hot aisle, rises through a grille into the ceiling return plenum and returns to the top of the CRAH. The dashed red arrow over the rack is recirculation over the rack top. Every element on the loop, the filter, the coil, the plenum, the tiles, the rack, the grille, is a resistance the fans have to push against.

6.3.2 Pressure and the quadratic law

Air flows from high pressure to low pressure, and the flow through an opening or a duct is set by the pressure difference across it. For the turbulent flows of a data hall the pressure drop across a fixed element is, to good approximation, proportional to the square of the flow through it,

$$\Delta p = K\,\dot{V}^2, \qquad \text{equivalently} \qquad \dot{V} = \sqrt{\Delta p / K} \;\propto\; \sqrt{\Delta p}, \tag{6.9}$$

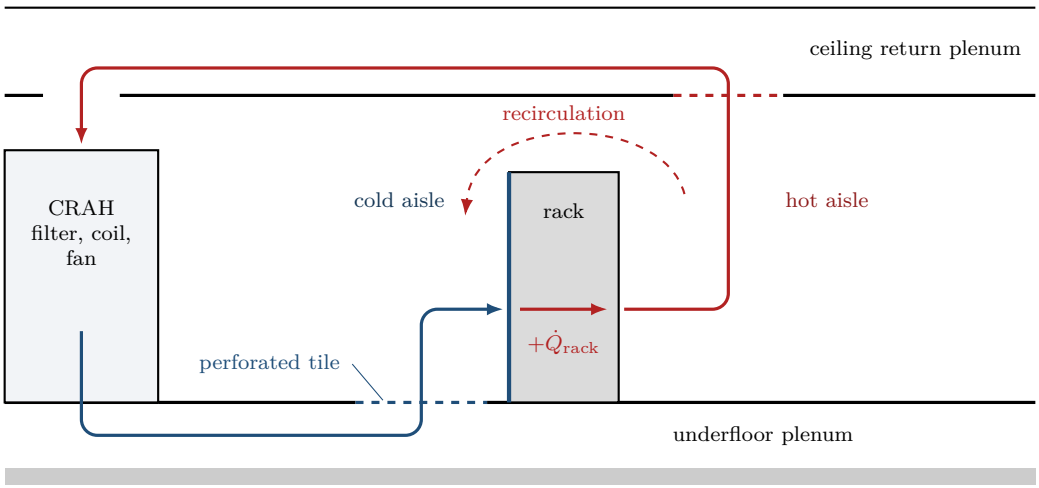


Figure 6.3: Section through a raised-floor hall, schematic. Supply air (blue) is pushed by the CRAH fan into the underfloor plenum, rises through perforated tiles (dashed floor segment) into the cold aisle and enters the rack front (thick blue edge). Heated air (red) leaves into the hot aisle, passes a ceiling grille (dashed ceiling segment) and returns through the ceiling plenum to the CRAH. The dashed arrow over the rack is recirculation.

where $\dot{V}$ is the volumetric flow rate in $\mathrm{m^3/s}$, Δp is the pressure difference in pascals and K is a resistance coefficient of the element. The square comes from the kinetic energy of the air: an opening of area A passing flow $\dot{V}$ has velocity $v = \dot{V}/A$, and the pressure lost is a multiple ζ of the dynamic pressure $\rho v^2/2$, so $K = \zeta\rho/(2A^2)$ with ρ the air density. The same law, with the same exponent, describes turbulent water in pipes (the Darcy–Weisbach relation), which is why air distribution and hydraulic networks are analyzed with the same tools.

The square root has a practical consequence. An illustrative perforated tile passes $0.25\,\mathrm{m^3/s}$ at a plenum pressure 10 Pa above the room, so $K = 10/0.25^2 = 160\,\mathrm{Pa\,s^2/m^6}$. Doubling the plenum pressure to 20 Pa raises the flow only to $\sqrt{20/160} = 0.354\,\mathrm{m^3/s}$, 41 % more, not 100 % more; doubling the flow would take four times the pressure. Pressure is an expensive way to buy flow, and a rack that is starved because its tile is too restrictive is not cured by a modest increase in plenum pressure.

6.3.3 Sources of resistance

Every element on the air loop takes a share of the pressure: the filters and the coil inside the cooling unit, the dampers and any ducts, the cables and pipes that obstruct the underfloor plenum, the perforated tiles, the rack doors and fronts, the server chassis and heat sinks, and the return grilles. Elements in series carry the same flow, so their pressure drops add,

$$\Delta p_{\text{total}} = \sum_i \Delta p_i = \Big(\sum_i K_i\Big)\,\dot{V}^2, \tag{6.10}$$

and the loop behaves as one resistance whose coefficient is the sum of the parts. (Elements in parallel, such as the tiles of one aisle, share the same pressure difference instead, and their flows add.)

For an illustrative CRAH loop, take a total coefficient of $2.0\,\mathrm{Pa\,s^2/m^6}$, made up of 0.5 for the filter, 0.6 for the coil, 0.5 for the plenum and tiles and 0.4 for the racks and the return. At $10\,\mathrm{m^3/s}$ the loop drops $2.0 \times 100 = 200\,\mathrm{Pa}$, of which the filter takes 50 Pa. As the filter collects dust its coefficient rises; at 2.5 the loop coefficient becomes 4.0. If the fan held the same 200 Pa, the flow would fall to $\sqrt{200/4} = 7.07\,\mathrm{m^3/s}$, a 29 % reduction. Cooling to every rack on that

unit falls, with no server failure, no alarm from the IT equipment and no change in the cooling plant: a maintenance item has become a thermal problem.

6.3.4 Fan curves, system curves and the operating point

A fan does not hold a fixed pressure. At a given speed it produces a pressure rise that falls as the flow it delivers increases; that relation, $\Delta p_{\text{fan}}(\dot{V})$, is the fan curve, supplied by the manufacturer. The loop it pushes against has its own relation, the system curve of equation (6.10), $\Delta p_{\text{system}} = K\dot{V}^2$. The fan delivers the flow at which the two agree,

$$\Delta p_{\text{fan}}(\dot{V}) = \Delta p_{\text{system}}(\dot{V}) = K\,\dot{V}^2, \tag{6.11}$$

the operating point. It is the same construction as a pump curve meeting the system curve of a hydraulic circuit, and the same reasoning applies: neither the fan alone nor the system alone determines the flow.

Take an illustrative fan curve $\Delta p_{\text{fan}} = 600 - 4\dot{V}^2$ (pascals, with $\dot{V}$ in m^3/s): 600 Pa at zero flow, falling to zero at about $12.2\,\text{m}^3/\text{s}$. With the clean loop, $K = 2$, equation (6.11) gives $600 - 4\dot{V}^2 = 2\dot{V}^2$, so $\dot{V}^2 = 100$, $\dot{V} = 10\,\text{m}^3/\text{s}$ and $\Delta p = 200\,\text{Pa}$. That is operating point A in Figure 6.4. The horizontal axis of the figure is the volumetric flow and the vertical axis the pressure; the solid descending curve is the fan curve, the dashed blue parabola the clean system curve and the dash-dotted red parabola the system curve with a clogged filter, treated next. Each operating point is an intersection of the fan curve with one system curve.

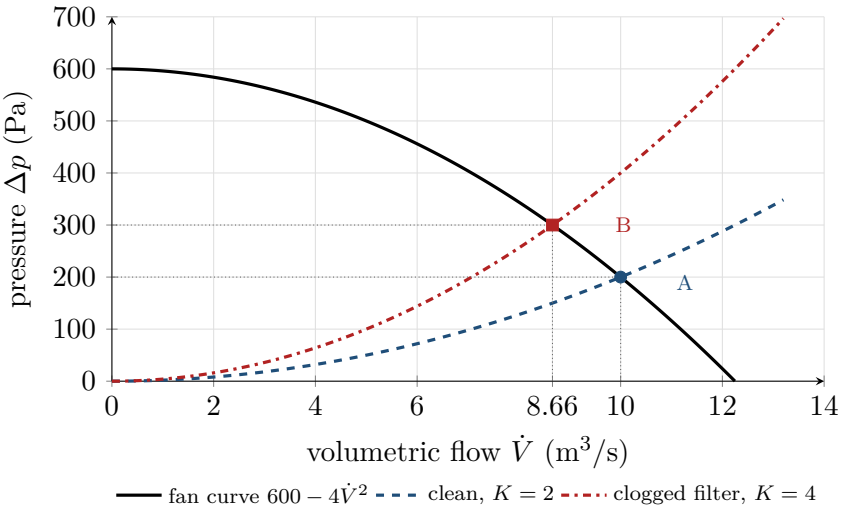


Figure 6.4: Fan curve and system curves, illustrative. Fan: $\Delta p_{\text{fan}} = 600 - 4\dot{V}^2$. Systems: $\Delta p = K\dot{V}^2$ with $K = 2$ (clean) and $K = 4$ (clogged filter), coefficients in $\text{Pa}\,\text{s}^2/\text{m}^6$. Operating point A (circle): $10\,\text{m}^3/\text{s}$ at 200 Pa; operating point B (square): $8.66\,\text{m}^3/\text{s}$ at 300 Pa, at the same fan speed.

6.3.5 The clogged filter and its diagnostic ambiguity

When the filter clogs, K rises from 2 to 4, the system curve becomes steeper, and the operating point slides up and to the left along the unchanged fan curve: $600 - 4\dot{V}^2 = 4\dot{V}^2$ gives $\dot{V} = \sqrt{75} = 8.66\,\text{m}^3/\text{s}$ at 300 Pa, point B of Figure 6.4. The flow falls by 13 %, less than the 29 % of the fixed-pressure calculation of Section 6.3.3, because the fan, meeting more resistance, produces more pressure. The fan speed is unchanged, and a monitoring system that watches only fan speed and fan status sees nothing. The racks see 13 % less air.

A lower flow at an unchanged fan speed command has several possible causes, and the flow alone does not distinguish them: the fan actually running slower than commanded (a drive or belt problem), a blocked filter, an obstructed rack or underfloor path, a damper in the wrong position, a change in room pressure (for example a door or a neighboring unit's operation), or a degraded fan whose curve has dropped. Some produce the same reduced flow through different curves: the blocked filter and the obstructed path steepen the system curve, while the slow and the degraded fan lower the fan curve. Separating them needs measurements placed on the loop, a pressure difference across the filter, the fan's measured speed and electrical power, a pressure in the plenum, rather than more measurements of the symptom. This is the first appearance of a pattern that recurs through the book: one observed symptom, several hypotheses, and a choice of which additional measurement discriminates them.

6.3.6 The affinity laws and the cost of lower speed

For a given fan in a given system, changing the rotational speed N_{fan} scales the operating point in a fixed way. Flow scales with speed, pressure with its square, and power with its cube,

$$\dot{V} \propto N_{\text{fan}}, \qquad \Delta p \propto N_{\text{fan}}^2, \qquad P_{\text{fan}} \propto N_{\text{fan}}^3. \tag{6.12}$$

The first two follow from the fan's geometry: each revolution moves the same volume, and the pressure a rotor imparts scales with the square of its tip speed. The third follows from the first two, because the power delivered to the air is the product $\Delta p \dot{V}$ divided by the fan efficiency. Because the system curve $K\dot{V}^2$ is itself a parabola through the origin, the laws hold along it: in the illustrative system of Figure 6.4, at 80 % speed the fan curve becomes $0.64 \times 600 - 4\dot{V}^2$, and with $K = 2$ the operating point is 8 m³/s at 128 Pa, exactly 0.8 times the flow and 0.64 times the pressure of point A.

At 80 % speed the power is $0.8^3 = 0.512$ of full-speed power, about 51 %: a 20 % reduction in flow nearly halves the fan power. A fan drawing 10 kW at full speed draws about 5.1 kW at 80 % speed. The same cube makes two fans at half speed draw $2 \times 0.5^3 = 0.25$ of the power of one fan at full speed for the same total flow, which is why facilities run all available units at reduced speed rather than a few at full speed. Variable-speed drives on facility fans are among the largest efficiency measures in a data hall. The cube is an upper bound on the saving: when the system must hold a minimum pressure regardless of flow, for example a plenum pressure setpoint, its curve does not pass through the origin and the power falls more slowly than N_{fan}^3.

Figure 6.5 plots the three laws. The horizontal axis is the fan speed as a fraction of full speed; the vertical axis is each quantity as a fraction of its full-speed value. The dotted guides mark the 80 % point, where the flow is 0.8, the pressure 0.64 and the power 0.512. The power curve lies far below the flow curve over the whole range: modest reductions in speed buy large reductions in power.

The cost of lower speed. The saving is not free. Lower speed means lower flow, and with the heat unchanged, equation (6.1) requires a larger temperature rise. If the 30 kW rack of Section 6.1.2 receives 80 % of its 2.5 kg/s, that is 2.0 kg/s, its rise grows from 12 K to $30/2.0 = 15$ K; if the reduced supply no longer covers what the server fans draw, recirculation makes up the difference and the inlet temperature rises too. Fan energy is traded against thermal margin, the distance between the hottest inlet and its limit. Choosing the speed that spends just enough margin, and no more, for the present loads is a control problem, formulated in Section 6.6.5.

6.3.7 Server fans and facility fans

Two sets of fans move the air. The facility fans, in the CRAH or computer room air conditioner (CRAC) units, deliver air to the room and pressurize the supply path. The server fans, inside

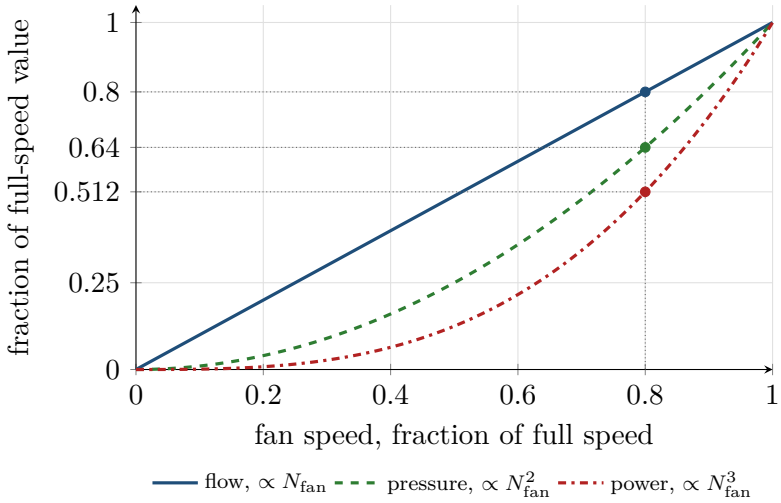


Figure 6.5: Fan affinity laws, equation (6.12), as fractions of full-speed values: flow $= n$, pressure $= n^2$, power $= n^3$, with n the speed fraction. At $n = 0.8$ the fan delivers 80 % of the flow for $0.8^3 = 0.512$ of the power.

every chassis, pull air through the chassis from the inlet to the exhaust. The two are in series on the same loop, and they interact. When the room delivers too little air or too little pressure in front of a rack, the servers' inlet and component temperatures rise, their fan controllers respond by raising the fan speed, and the server fans pull harder, drawing whatever air is available, including recirculated exhaust.

Server fans obey the same cube law, and their power is part of the IT load. An illustrative rack of 40 servers whose fans draw 1.2 kW in total at their normal speed draws $1.3^3 \times 1.2 = 2.6$ kW when a starved inlet drives them 30 % faster: 1.4 kW more electrical power, which becomes 1.4 kW more heat for the same computation. The chain

$$\text{airflow problem} \rightarrow \text{server fan speed} \uparrow \rightarrow \text{server fan power} \uparrow \rightarrow \text{IT power and heat} \uparrow$$

means that an airflow deficiency increases the heat it was failing to remove, and it shows up in the electrical measurements as a higher IT load. Because that power is counted as IT power, it also flatters the facility's efficiency ratio, a point Chapter 11 returns to; the interaction of server fan control with utilization is part of the server power model of Chapter 2.

6.4 Distribution, containment and hot spots

The previous sections dealt with one rack and one fan loop. This section scales up to a hall: the balance of total airflow, the different needs of different racks, the distinction between a capacity problem and a distribution problem, containment as the main remedy for mixing, and the diagnosis of a hot spot.

6.4.1 Airflow mismatch in a hall

Air is conserved. If the racks of an illustrative hall draw 100 000 m^3/h through their server fans and the cooling units supply 85 000 m^3/h, the racks are short by

$$100\,000 - 85\,000 = 15\,000\,\text{m}^3/\text{h},$$

15 % of what they draw. The server fans do not stop at the end of the supply; they pull the remaining 15 000 m^3/h from the only other source in the room, the hot exhaust. The hall-wide

recirculation fraction is therefore at least 0.15, and more if any of the supply bypasses. In mass terms, at 1.2 kg/m^3, the racks draw about 33.3 kg/s, the supply delivers 28.3 kg/s and the deficit is 5.0 kg/s.

The temperature consequence follows from equation (6.7). If the recirculated air came evenly from racks with a 12 K rise, every inlet would sit $(0.15/0.85) \times 12 = 2.1$ K above the supply, which is tolerable. It does not come evenly: the deficit concentrates at the racks farthest from the cooling units and nearest the openings between aisles, where the local recirculation fraction can be several times the hall average. A hall-wide airflow balance is a necessary check, and a deficit in it is a certain sign of recirculation, but its average says little about the worst rack.

6.4.2 Rack airflow demand

Each rack's need follows from equation (6.2) applied rack by rack,

$$\dot{m}_i = \frac{\dot{Q}_i}{c_{p,\mathrm{air}}\,(T_{\mathrm{out},i} - T_{\mathrm{in},i})}, \tag{6.13}$$

and racks differ widely. At a 12 K rise, a 10 kW rack needs $10/12 = 0.83$ kg/s and an 80 kW rack needs $80/12 = 6.67$ kg/s, eight times as much, about 5.6 m^3/s or 20 000 m^3/h. With the illustrative tile of Section 6.3.2, which passes 0.25 m^3/s at 10 Pa, the 10 kW rack is served by about three tiles and the 80 kW rack would need more than twenty, in front of a rack that occupies the width of one or two. Air delivered through a floor cannot be concentrated that much in one place, which is one reason the highest-density racks move to liquid cooling, treated in Chapter 7.

High-density racks placed in an older hall designed for an even, modest density produce exactly this mismatch: the cooling units together have enough capacity for the total heat, but the floor, the tiles and the plenum cannot deliver the concentrated airflow the new racks demand, while the old racks around them are oversupplied.

6.4.3 Capacity problems and distribution problems

Two different problems produce the same symptom, a hot inlet. A *capacity* problem is a shortage of total cooling: the sum of available cooling is below the sum of the heat, and no rearrangement of the air can fix it. A *distribution* problem is a shortage at particular racks while the total is sufficient: $\dot{Q}_{\mathrm{avail}} \geq \dot{Q}_{\mathrm{req}}$ holds for the room, and $\dot{Q}_{\mathrm{avail},i} < \dot{Q}_{\mathrm{rack},i}$ for some racks, as in equation (6.3).

The remedies differ, and confusing the two is costly. A capacity problem calls for more cooling equipment or less load. A distribution problem calls for moving tiles, sealing cable openings, installing blanking panels in empty rack positions, closing gaps between racks, adding containment or redistributing load; it is not solved by a new cooling unit, whose air enters the same distorted distribution, and a new unit added to a hall with bypass increases the bypass. The airflow balance of Section 6.4.1, the per-rack demands of Section 6.4.2 and the per-rack inlet temperatures of Section 6.2.6 together distinguish the two.

6.4.4 Containment

Containment closes the paths along which supply and exhaust mix. *Cold-aisle containment* encloses the cold aisle with doors at its ends and a roof over it, so that supply air can leave only through the rack inlets; the rest of the room becomes part of the hot return. *Hot-aisle containment* encloses the hot aisle and connects it to the return path, often through a ceiling plenum, so that exhaust can go only back to the cooling units; the rest of the room becomes part of the cool supply. Blanking panels in empty rack positions and brush seals in floor cable openings are smaller forms of the same idea.

Both forms reduce both failures. With the cold aisle closed, exhaust cannot reach the inlets around the ends of rows or over the rack tops, so α falls, and supply air cannot leave the aisle except through the racks, so β falls. Hot-aisle containment achieves the same from the other side. In the plan view of Figure 6.1, containment removes both dashed paths. Neither form is perfect: leakage through gaps between racks and through the racks' own structure leaves a small residual recirculation and bypass, which is why measured temperatures in a contained hall still differ from rack to rack.

6.4.5 Why containment saves energy

The saving does not come from the enclosure itself but from what it allows. With mixing, the supply temperature must be set low enough that the worst rack, with the highest α, still meets its limit. With less mixing, the supply can be warmer for the same worst inlet. Take the illustrative limit $T_{\max} = 27°\mathrm{C}$ and a hot aisle 18 K above the supply. Before containment the worst rack has $\alpha = 0.25$, so its inlet is $T_{\text{supply}} + 0.25 \times 18 = T_{\text{supply}} + 4.5\,\mathrm{K}$ and the supply can be at most 22.5°C. After containment the worst rack has $\alpha = 0.05$, its inlet is $T_{\text{supply}} + 0.9\,\mathrm{K}$, and the supply can rise to 26.1°C, 3.6 K warmer.

A warmer supply lets the chilled water run warmer, and a chiller makes cold water more efficiently when the water is less cold; a commonly quoted illustrative figure is a few percent of chiller energy per kelvin. A warmer supply also extends the hours in which outdoor air or a cooling tower alone can do the work. Reduced bypass adds a fan saving: to put 75 kg/s through the IT equipment, a hall with $\beta = 0.25$ must supply 100 kg/s, and one with $\beta = 0.05$ only $75/0.95 = 79\,\mathrm{kg/s}$, which by the cube law of equation (6.12) takes $0.79^3 \approx 0.49$ of the fan power. Energy is saved with no change in the IT load, by changing the paths of the air. Chapter 7 develops the dependence of the cooling plant on these temperatures.

Principle 6.3. *Containment saves energy indirectly: by reducing recirculation and bypass, it lets the supply air and chilled water run warmer and the fans run slower for the same worst rack inlet temperature.*

6.4.6 Hot spots

A hot spot is a rack, or a part of a rack, whose inlet temperature is well above its neighbors'. Three mechanisms produce one, alone or together:

- a higher heat load, P_{rack} rising, which raises the exhaust and, through any recirculation, the inlet;
- a lower rack airflow, $\dot{m}_{\text{rack}}$ falling, from a blocked tile, an obstruction or a starved end of a row;
- a higher recirculation fraction, α rising, from a missing blanking panel, an open containment door or a pressure imbalance.

The rising inlet temperature $T_{\text{rack,in}}$ is therefore a symptom, not a cause. Treating it as a cause, for instance by lowering the supply temperature of the whole hall, cools every rack to correct one and leaves the mechanism in place.

The three mechanisms leave different traces in other measurements. A higher load appears in the rack's electrical power. A lower airflow shows as a larger rise across the rack for the same power and often as a lower pressure difference between the cold and hot aisles at that rack. A higher recirculation fraction shows as a warmer hot aisle next to the rack, or a temperature gradient up the rack front, warmest at the top where exhaust spills over. The diagnosis uses these traces.

6.4.7 A diagnostic example: rack 18

In an illustrative hall, rack 18 reports an inlet temperature of 29°C while its neighbors read 23°C. With the supply at 20°C and the adjacent hot aisle at 38°C, equation (6.6) would attribute the difference to a recirculation fraction of $(29 - 20)/18 = 0.5$ at rack 18 against $3/18 \approx 0.17$ at its neighbors, but that is the reading under one hypothesis among several. Five candidate explanations are listed in Table 6.1: an unusually high load (H_1), a local airflow obstruction (H_2), recirculation of hot exhaust (H_3), a blocked perforated tile (H_4), and a biased inlet sensor (H_5).

The table lists, for each hypothesis, what it predicts in the other available measurements and what the evidence in this example says. The rack's electrical power is normal, which makes H_1 much less plausible: a high-load hot spot must show in the power. The measured pressure difference between the cold aisle and the hot aisle at rack 18 is low, which supports an airflow problem (H_2, H_3 or H_4) and weighs against a pure sensor fault. The adjacent hot aisle is warmer than elsewhere in the row, which makes recirculation (H_3) more plausible. No single reading decides the case. Each one shifts the relative plausibility of the hypotheses, and the result is a ranking: here H_3 first, H_2 and H_4 next, H_5 and H_1 last, with the probability of each a quantity that the measurements update rather than a verdict. Section 6.6.6 asks which measurement to add next.

Table 6.1: Hypotheses for the 29°C inlet at rack 18 (neighbors 23°C), illustrative: what each predicts and how the available evidence bears on it.

Hypothesis	Predicts	Evidence in the example
H_1 unusually high load	rack power above normal; larger rise across the rack	power normal: strongly weakened
H_2 local airflow obstruction	low rack airflow; low aisle pressure difference at the rack; large rise across the rack	low pressure difference: supported
H_3 hot-exhaust recirculation	warmer adjacent hot aisle; inlet warmest at the top of the rack front	low pressure difference and warm hot aisle: most supported
H_4 blocked perforated tile	low tile flow and low local cold-aisle pressure; neighbors on the same tile row affected	low pressure difference: supported
H_5 inlet sensor bias	no change in rise, power or pressure; other sensors on the rack disagree	low pressure difference is not explained: weakened

6.5 Measuring and inferring airflow

The quantity that decides local adequacy, the airflow through each rack, is the one a data hall rarely measures. This section describes the pressure measurements that are available, explains why rack airflow is inferred rather than measured, works the inference through an example, and states it as a probability.

6.5.1 Static pressure as evidence

Halls commonly measure static pressure: the pressure of the air at rest relative to a reference, in the underfloor plenum and between the cold and hot aisles. The aisle pressure difference

$$\Delta p = p_{\text{cold aisle}} - p_{\text{hot aisle}} \tag{6.14}$$

indicates the direction of the leakage flows. A positive difference of a few pascals (an illustrative order of magnitude) pushes air from the cold side to the hot side through every gap, so that leaks carry supply air into the hot aisle rather than exhaust into the cold aisle. A difference that is too low, or negative, lets exhaust flow back through those gaps into the cold aisle: recirculation. Controllers in contained halls often regulate facility fan speed to hold this difference at a small positive setpoint.

Pressure is indirect evidence of airflow. By equation (6.9), a pressure difference determines a flow only through a resistance, $\dot{V} = \sqrt{\Delta p / K}$, and K for a rack is not known: it depends on the rack's doors, cabling and contents, and the rack is not a passive element at all, because its server fans add their own pressure. A normal aisle pressure difference is consistent with a starved rack whose fans are pulling hard, and a low one is consistent with a rack drawing its full flow from a neighbor's tiles. Pressure narrows the possibilities; without the resistance relation it does not give the mass flow, $\Delta p \neq \dot{m}$.

6.5.2 Why rack airflow is poorly measured

The measurements commonly available at a rack are its inlet temperature, often its exhaust temperature, and its electrical power from the power distribution unit. Its airflow is rarely measured. Air velocity varies strongly across the face of a rack, from server to server and from top to bottom, so a point anemometer reads one location, not the rack's flow; flow hoods and velocity grids that integrate over the face are intrusive, portable and used in surveys, not left in place. A rack's mass flow is therefore estimated from the energy balance, equation (6.1), solved for the flow,

$$\dot{m}_{\text{air}} \approx \frac{P_{\text{rack}}}{c_{p,\text{air}}\,(T_{\text{out}} - T_{\text{in}})}, \tag{6.15}$$

with the measured electrical power standing in for the heat.

In the book's three categories, P_{rack}, T_{in} and T_{out} are measured, and $\dot{m}_{\text{air}}$ is neither measured nor simply calculated: it is an inferred state, obtained from measurements through a model whose assumptions may fail. The difference matters because the assumptions, not the arithmetic, are where the estimate goes wrong.

6.5.3 An inferred airflow of 2.0 kg/s

An illustrative rack reports $P_{\text{rack}} = 30\,\text{kW}$, $T_{\text{in}} = 22°\text{C}$ and $T_{\text{out}} = 37°\text{C}$. The rise is $\Delta T = 15\,\text{K}$ and equation (6.15) gives

$$\dot{m}_{\text{air}} \approx \frac{30\,\text{kW}}{1.0\,\text{kJ/(kg K)} \times 15\,\text{K}} = 2.0\,\text{kg/s},$$

without any airflow sensor. The estimate rests on four assumptions, each with a known direction of error.

(a) *All electrical power becomes sensible heat in the air.* Very nearly true for IT equipment, but heat that leaves by another path, such as a liquid-cooled component or a rear-door heat exchanger, makes the air's heat smaller than P_{rack} and the estimate too high.

(b) *The temperatures represent the bulk inlet and outlet streams.* One exhaust sensor near the top of a rack, where the exhaust is often hottest, overstates the mean rise and makes the estimate too low; an inlet sensor in a cooler spot than the average inlet has the opposite effect.

(c) *Humidity effects are small.* The specific heat of moist air differs from that of dry air by one to two percent at room conditions, and there is no condensation or evaporation inside a server, so this assumption is usually safe.

(d) *No other significant heat path.* Conduction through the rack walls and radiation to the room are small for a rack in a moving air stream, but not zero.

The estimate also inherits the measurement errors. A 1 K error in the rise changes it from $30/15 = 2.0$ to $30/16 = 1.875$ kg/s or $30/14 = 2.14$ kg/s, a change of about 7 %. Because the rise is in the denominator, the error grows as the rise shrinks: the same 1 K error on a 5 K rise is a 20 % error in the flow.

6.5.4 A probabilistic version

The honest statement of the estimate is a distribution, not a number:

$$p(\dot{m}_{\text{air}} \mid P_{\text{rack}}, T_{\text{in}}, T_{\text{out}}), \tag{6.16}$$

the probability density of the rack airflow given the three measurements, each of which is uncertain. If the power is measured with a standard deviation of 2 % and each temperature with 0.5 K, the rise has a standard deviation of $\sqrt{0.5^2 + 0.5^2} = 0.71$ K, or 4.7 % of 15 K, and to first order the relative uncertainties of a ratio combine in quadrature,

$$\frac{\sigma_{\dot{m}}}{\dot{m}_{\text{air}}} \approx \sqrt{0.02^2 + 0.047^2} \approx 0.051,$$

so the airflow is about 2.0 ± 0.10 kg/s (one standard deviation). The distribution is not symmetric: because the rise sits in the denominator, its upper tail is longer than its lower tail, and the asymmetry grows as the rise shrinks.

The bulk-temperature and heat-path assumptions of Section 6.5.3 enter the same statement as biases with their own uncertainty: an exhaust sensor known to read high relative to the mixed exhaust becomes an offset with a distribution, rather than an unmodeled error. Written this way, the airflow estimate is a posterior distribution, $p(\dot{m}_{\text{air}} \mid y) \propto p(y \mid \dot{m}_{\text{air}})\, p(\dot{m}_{\text{air}})$, with y the measurements, a likelihood built from equation (6.1) and the measurement errors, and a prior that can carry what is known about the rack's design flow. The same construction applies to every inferred quantity of the chapter: a recirculation fraction, a resistance coefficient or a fan's degradation.

> **Principle 6.4.** *Rack airflow is an inferred state. It is obtained from measured power and temperatures through an energy balance whose assumptions may fail, and it should be stated as a distribution whose width is set by the measurement errors and by the size of the temperature rise.*

6.6 Network models, CFD and control

The relations of the previous sections, a balance per rack, a mixing law per inlet, a pressure law per path, fit together into a network model of the whole hall. This section builds that model, compares it with computational fluid dynamics, and uses it for the two decisions an operator faces: how to run the fans, and which measurement to add next.

6.6.1 The airflow network

A reduced model of a data hall represents the air volumes as nodes and the paths between them as edges. The nodes are the supply plenum, the cold aisles, the racks, the hot aisles and the return plenum; each is assumed well mixed, with one pressure and one temperature. The edges are the airflow paths: tiles from the plenum to a cold aisle, racks from a cold aisle to a hot aisle, grilles from a hot aisle to the return, the cooling unit from the return back to the supply, and the leakage paths, through gaps and over rack tops, that carry recirculation and bypass. Each edge e carries a mass flow set by the pressure difference across it and its resistance,

$$\dot{m}_e = f(\Delta p_e, K_e) = \rho \, \mathrm{sign}(\Delta p_e) \sqrt{|\Delta p_e| / K_e}, \tag{6.17}$$

the quadratic law of equation (6.9) written for either direction of flow. Fans are edges that add pressure according to their curves. At every node mass is conserved, $\sum \dot{m}_{\mathrm{in}} = \sum \dot{m}_{\mathrm{out}}$.

Figure 6.6 draws such a network for a small hall. Shaded rounded boxes are air-volume nodes (blue outline on the supply side, red on the return side) and white boxes are racks; arrows are edges, blue for supply air and red for exhaust. The supply plenum feeds two cold aisles through their tiles; each cold aisle feeds two racks, whose heat inputs $\dot{Q}_i$ are marked; all four racks exhaust into one hot aisle, which returns through grilles to the return plenum and the CRAH, whose fan closes the loop. The dashed red edge from the hot aisle back to cold aisle 1 is recirculation; the dashed blue edge from cold aisle 2 to the return plenum is bypass. The structure is that of hydraulic networks, gas networks and electrical circuits: node balances play the role of Kirchhoff's current law and pressures the role of voltages, with a quadratic instead of a linear relation on each edge. Given the fan speeds and the resistances, the node pressures and edge flows follow from solving the mass balances, a set of nonlinear equations of the same kind as a power-flow or pipe-network problem.

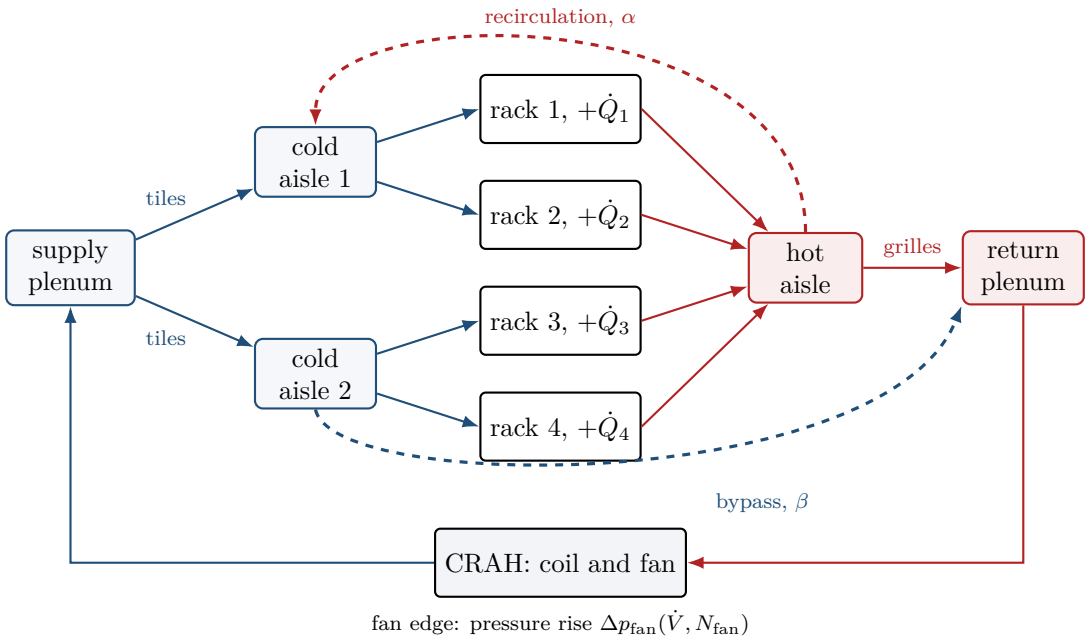


Figure 6.6: Airflow network of a small hall, schematic. Nodes are air volumes (supply plenum, cold aisles, hot aisle, return plenum) and racks, each assumed well mixed; edges are airflow paths with $\dot{m}_e = f(\Delta p_e, K_e)$ (equation (6.17)). Mass is conserved at every node, and energy with the rack heat inputs $\dot{Q}_i$. Dashed edges: recirculation (red) and bypass (blue).

6.6.2 Mass and energy at a node

At a node i with inflows $\dot{m}_{ji}$ from nodes j and outflows $\dot{m}_{ik}$ to nodes k, the steady mass balance is

$$\sum_j \dot{m}_{ji} = \sum_k \dot{m}_{ik}, \tag{6.18}$$

and the steady energy balance, with each inflow carrying the temperature T_j of the node it comes from, each outflow leaving at the node's own temperature T_i (the well-mixed assumption) and $\dot{Q}_i$ the heat added at the node (zero except at racks), is

$$\sum_j \dot{m}_{ji}\, c_{p,\mathrm{air}}\, T_j + \dot{Q}_i = \sum_k \dot{m}_{ik}\, c_{p,\mathrm{air}}\, T_i. \tag{6.19}$$

The recirculation law, equation (6.5), is equation (6.19) at a cold-aisle node with two inflows, and the rack balance, equation (6.1), is the same equation at a rack node with one inflow and one outflow.

The 30 kW rack of Section 6.1.2 illustrates the pair. At the cold-aisle node in front of it, 1.8 kg/s of supply at 20°C and 0.7 kg/s of recirculated exhaust at 36.7°C enter, and 2.5 kg/s leaves at $(1.8 \times 20 + 0.7 \times 36.7)/2.5 = 24.7$°C. At the rack node, 2.5 kg/s at 24.7°C plus 30 kW leaves at $24.7 + 30/2.5 = 36.7$°C, which is the temperature assumed for the recirculated stream: the two balances close together. In general the model links pressure, airflow, temperature and heat load in one set of equations. Once the flows are known, the energy balances are linear in the temperatures, so a reduced model of a whole data hall solves in a fraction of a second without resolving the flow field in detail.

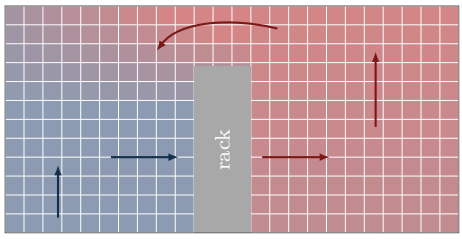
6.6.3 Computational fluid dynamics and the reduced model

Real halls are not a handful of well-mixed volumes. They contain racks of different heights, cable trays, walls, columns, floor openings, turbulent jets from tiles and diffusers, wakes behind obstacles, and buoyant plumes of warm air that rise and spread along the ceiling. Computational fluid dynamics resolves all of this. It divides the room into a grid of small cells and solves, in each cell, the conservation of mass, which for the nearly incompressible air of a data hall is $\nabla \cdot \mathbf{u} = 0$ for the velocity field $\mathbf{u}$, together with the momentum equations (the Navier–Stokes equations with a turbulence model and a buoyancy term) and the energy equation for the temperature field. The result is a detailed temperature and velocity field throughout the room. Its price is computation: millions of cells, each with several unknowns, solved iteratively, so one steady case takes minutes to hours, and the model needs detailed geometry, boundary conditions and calibration.

The two kinds of model serve different purposes. Figure 6.7 sets them side by side for the same cold aisle, rack and hot aisle. The left panel represents a CFD solution: a grid of cells over a vertical section, shaded from cool (blue) to warm (red), with the rack as a gray block and a warm plume curling over the rack top into the cold aisle. The right panel is the reduced network for the same space: five nodes and the edges between them, including a recirculation edge whose fraction α stands for the whole plume. Below each panel are the model's typical unknowns, computing time and uses, as orders of magnitude. CFD is the tool for detailed design, hot-spot analysis and airflow visualization; it is expensive, slow and parameter intensive. The reduced network is the tool for real-time estimation, repeated what-if calculations, probabilistic analysis and control; it gives up spatial detail. The open question that the reduced model has to answer for each hall is whether a physical model this coarse captures enough of the structure, the right nodes, the right paths, parameters that can be inferred from the available measurements, to support diagnosis and risk analysis at a small fraction of the cost of repeated CFD runs. One practical answer combines the two: CFD runs once, offline, to suggest the network's paths

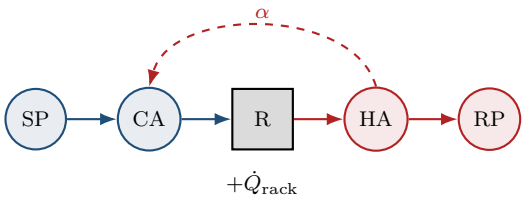
and the plausible ranges of its parameters, and the network runs continuously against the measurements.

CFD: a field on a grid



unknowns: velocity, pressure and temperature in each of 10^6–10^8 cells
one steady case: minutes to hours
needs: detailed geometry, boundary conditions
uses: detailed design, hot-spot analysis, airflow visualization

reduced network



unknowns: one pressure and one temperature per node, tens to thousands
one steady case: milliseconds to seconds
needs: resistances, fan curves, mixing fractions (inferred)
uses: real-time estimation, what-if, probabilistic analysis, control

Figure 6.7: The same cold aisle (left of the rack) and hot aisle (right) in two models, schematic. Left: a CFD solution as a shaded temperature field on a grid, with a warm plume over the rack top. Right: the reduced network (SP supply plenum, CA cold aisle, R rack, HA hot aisle, RP return plenum), in which the plume is one recirculation edge with fraction α. Unknown counts and times are typical orders of magnitude, illustrative.

6.6.4 Identical racks with different risk

Two illustrative racks draw the same electrical power, $P_1 = P_2 = 40\,\mathrm{kW}$. Rack 1 receives $\dot{m}_1 = 4\,\mathrm{kg/s}$ and rack 2 receives $\dot{m}_2 = 2\,\mathrm{kg/s}$. By equation (6.1) their rises are

$$\Delta T_1 = \frac{40}{4 \times 1.0} = 10\,\mathrm{K}, \qquad \Delta T_2 = \frac{40}{2 \times 1.0} = 20\,\mathrm{K}.$$

With a 22°C inlet, rack 1 exhausts at 32°C and rack 2 at 42°C, with correspondingly hotter components and faster server fans. If the hot aisle they share also feeds any recirculation, rack 2's exhaust raises the inlets downstream of it twice as much as rack 1's.

Figure 6.8 draws the relation that separates them. The horizontal axis is the airflow through a rack; the vertical axis is the temperature rise across it. The single curve is $\Delta T = 40\,\mathrm{kW}/(\dot{m}\, c_{p,\mathrm{air}})$, the same for both racks because their power is the same; the two dots are the racks' operating points. The dotted vertical line marks the flow, $40/15 = 2.67\,\mathrm{kg/s}$, below which the rise exceeds an illustrative allowable 15 K. The curve is steep at low flow, so a rack already short of air loses margin quickly as its flow falls further. A power meter reports 40 kW for both racks and cannot tell them apart: thermal risk cannot be inferred from P_{rack} alone. It depends on the heat-removal path, the airflow and the mixing that bring air to the rack, which is what the network model represents.

6.6.5 The control problem

The trade-off of Section 6.3.6, fan energy against thermal margin, becomes an optimization once the network model predicts each rack's inlet temperature as a function of the control inputs u (fan speeds, supply temperature, damper or tile positions). In deterministic form,

$$\min_{u} \; P_{\mathrm{fan}}(u) \qquad \text{subject to} \qquad T_{\mathrm{in},i}(u) \le T_{\max} \quad \text{for every rack } i, \tag{6.20}$$

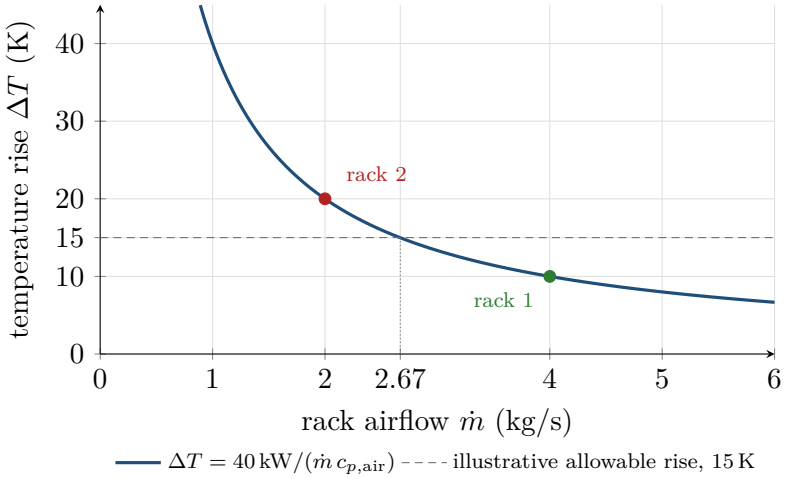


Figure 6.8: Temperature rise against airflow for two identical 40 kW racks, $\Delta T = P_{\text{rack}}/(\dot{m}\, c_{p,\text{air}})$ with $P_{\text{rack}} = 40\,\text{kW}$ and $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$. Rack 1 at 4 kg/s rises 10 K; rack 2 at 2 kg/s rises 20 K. The allowable rise is illustrative.

the controller looks for the least fan power that keeps every inlet, including the worst, within its limit. Because P_{fan} falls as the cube of speed, the optimum is usually at the lowest speed at which the worst rack just meets its limit.

The inlet temperatures are not known exactly. They depend on inferred quantities, recirculation fractions, rack airflows, sensor biases, whose uncertainty the previous section made explicit. The deterministic constraint is replaced by a chance constraint,

$$\Pr\big(T_{\text{in},i}(u) > T_{\max}\big) < \epsilon \quad \text{for every rack } i, \tag{6.21}$$

which requires each limit to hold with probability at least $1-\epsilon$: probabilistic control. If the predicted inlet of rack i is normally distributed with mean $\mu_i(u)$ and standard deviation σ_i, the chance constraint is equivalent to $\mu_i(u) + z_{1-\epsilon}\,\sigma_i \le T_{\max}$, with $z_{1-\epsilon}$ the standard normal quantile; for $\epsilon = 0.01$, $z = 2.33$.

An illustrative case shows the cost of uncertainty. Let the worst rack follow $T_{\text{in}} = 20 + 18\,\alpha(u)$ with $\alpha(u) = 1 - u/0.9$ for a fan speed fraction $u \le 0.9$ (the supply matches the rack's draw at 90 % speed and falls short below it), and $T_{\max} = 27°\text{C}$. The deterministic constraint gives $\alpha \le 0.389$, so $u \ge 0.55$, with fan power $0.55^3 = 0.17$ of full power. With $\sigma = 1\,\text{K}$ and $\epsilon = 0.01$ the mean inlet must stay below $27 - 2.33 = 24.67°\text{C}$, so $\alpha \le 0.26$, $u \ge 0.67$ and the fan power is 0.30 of full. With a better estimate, $\sigma = 0.5\,\text{K}$, the mean may rise to $25.8°\text{C}$, $u \ge 0.61$ and the power falls to 0.23. Uncertainty costs fan energy, and better inference returns it. Chapter 10 develops the controllers that implement such problems.

6.6.6 The next-best measurement

Return to rack 18 of Section 6.4.7 and suppose the evidence has narrowed the explanation to two hypotheses: low airflow through the rack, H_2 of Table 6.1 (taken here to include the blocked tile H_4, which acts the same way), and recirculation of hot exhaust, H_3. One additional measurement can be installed or taken by hand. The candidates are the rack airflow (a portable flow hood), the local static pressure in front of the rack, the hot-aisle temperature next to it, the server fan speeds, and the supply temperature. The best candidate is the one that most reduces the uncertainty between the hypotheses: the measurement with the largest value of information.

Table 6.2 makes that concrete with illustrative numbers. For each candidate it states what the measurement would show under each hypothesis, as the probability of an "abnormal" reading

(low flow, low pressure, warm hot aisle, high fan speed, off-setpoint supply), and the expected information the reading carries about which hypothesis holds. With equal prior probabilities, the information is the expected reduction in the uncertainty of the hypothesis, measured in bits: one bit would settle the question with certainty, zero leaves it where it was,

$$I(H; y) = \mathcal{H}(H) - \mathbb{E}_y[\mathcal{H}(H \mid y)], \tag{6.22}$$

with $\mathcal{H}$ the entropy of the distribution over the hypotheses. The rack airflow discriminates best, about 0.53 bits, because the two hypotheses predict opposite readings: an abnormal (low) reading raises the probability of H_2 from 0.5 to 0.9. The hot-aisle temperature helps less (0.19 bits) and the static pressure less still (0.07), because both hypotheses tend to lower it. The server fan speed carries almost nothing (0.01), because both hypotheses make the fans speed up, and the supply temperature carries nothing, because neither hypothesis changes it; it would be the right measurement only for a third hypothesis, a supply fault. The useful measurement is the one whose prediction differs between the hypotheses, not the one that is easiest to read or that responds most strongly to the symptom.

Table 6.2: Candidate next measurements for rack 18 under two hypotheses, H_2 low rack airflow and H_3 recirculation, with prior probability 0.5 each. Columns two and three: probability of an abnormal reading under each hypothesis (illustrative). Column four: expected information about the hypothesis, equation (6.22).

Measurement	abnormal $\mid H_2$	abnormal $\mid H_3$	information (bits)	Reason
Rack airflow	0.9	0.1	0.53	opposite predictions
Hot-aisle temperature	0.3	0.8	0.19	mostly a sign of H_3
Local static pressure	0.8	0.5	0.07	low under both
Server fan speed	0.8	0.7	0.01	high under both
Supply temperature	0.5	0.5	0.00	unaffected by either

6.6.7 The relations to keep and the delivery question

Five relations carry the chapter:

$$\dot{Q} = \dot{m}\, c_p\, \Delta T, \qquad \Delta p \approx K\, \dot{V}^2, \qquad T_{\text{in}} = (1-\alpha)\, T_{\text{supply}} + \alpha\, T_{\text{hot}},$$
$$\dot{V} \propto N_{\text{fan}}, \qquad \Delta p \propto N_{\text{fan}}^2, \qquad P_{\text{fan}} \propto N_{\text{fan}}^3. \tag{6.23}$$

The first is the energy balance of every rack and every node; the second is the resistance of every path; the third is the mixing at every inlet; the last three are the cost of moving the air. With the node balances of equations (6.18) and (6.19), they are a complete reduced model of a hall's air side.

They also restate the chapter's question. The question for a data hall is not only whether it has enough cooling capacity, but whether the required cooling can be transported to every rack that needs it. That second question is answered by airflow, pressure, resistance, mixing and distribution, the subjects of this chapter, and a hall can pass the first while failing the second at a single rack. Cooling capacity is not cooling delivery.

6.7 What to retain

Definition 6.1 (Cooling delivery)**.** A rack is cooled by the airflow that reaches its inlet and by that air's temperature, $\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}} c_{p,\text{air}} (T_{\text{out}} - T_{\text{in}})$ with $T_{\text{in}} = (1-\alpha) T_{\text{supply}} + \alpha T_{\text{hot}}$; adequate

cooling capacity for the room does not imply adequate delivery to every rack, and the delivery is set by pressures, resistances, fans and mixing, most of which are inferred rather than measured.

Table 6.3 lists the chapter's objects and the subsection that defines each.

Table 6.3: The objects of the airflow model and where each is defined.

Object	Expression	Role	Defined in
Rack airflow balance	$\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}} c_{p,\text{air}} (T_{\text{out}} - T_{\text{in}})$	heat carried by the air	6.1.1
Required airflow	$\dot{m}_{\text{air}} = \dot{Q}_{\text{rack}} / (c_{p,\text{air}} \Delta T)$	flow a rack needs	6.1.1
Local adequacy	$\dot{Q}_{\text{avail},i} \geq \dot{Q}_{\text{rack},i}$	the condition per rack	6.1.3
Recirculation fraction	α, equation (6.5)	mixing at a rack inlet	6.2.3
Bypass fraction	$\beta = \dot{m}_{\text{byp}} / \dot{m}_{\text{sup}}$	supply that does no cooling	6.2.4
Quadratic law	$\Delta p = K \dot{V}^2$	resistance of a path	6.3.2
Operating point	$\Delta p_{\text{fan}}(\dot{V}) = K \dot{V}^2$	flow a fan delivers	6.3.4
Affinity laws	$\dot{V} \propto N_{\text{fan}}$, $\Delta p \propto N_{\text{fan}}^2$, $P_{\text{fan}} \propto N_{\text{fan}}^3$	cost of moving air	6.3.6
Rack demand	$\dot{m}_i$, equation (6.13)	per-rack airflow need	6.4.2
Aisle pressure difference	$p_{\text{cold}} - p_{\text{hot}}$	indirect evidence of flow direction	6.5.1
Inferred airflow	$p(\dot{m}_{\text{air}} \mid P_{\text{rack}}, T_{\text{in}}, T_{\text{out}})$	airflow as a distribution	6.5.4
Node balances	equations (6.18)–(6.19)	the reduced network model	6.6.2
Chance-constrained control	equation (6.21)	fan energy against margin	6.6.5
Value of information	$I(H; y)$, equation (6.22)	choice of the next measurement	6.6.6

6.8 Exercises

Exercise 6.1. A rack draws 25 kW and is designed for an inlet of 21°C and an exhaust of 35°C. (a) Compute the required mass flow and, at 1.2 kg/m^3, the volumetric flow in m^3/s and m^3/h. (b) The tiles in front of the rack deliver only 1.5 kg/s of 21°C supply, and the server fans keep drawing the required flow, the rest coming from the rack's own exhaust. Compute the recirculation fraction and, with equation (6.7), the resulting inlet and exhaust temperatures.

Exercise 6.2. A CRAH fan has the curve $\Delta p_{\text{fan}} = 500 - 5\dot{V}^2$ (Pa, m^3/s) and pushes against a clean loop with $K = 3\,\text{Pa}\,\text{s}^2/\text{m}^6$. (a) Find the operating point. (b) The filter clogs and K rises to 5; find the new operating point and the percentage loss of flow at unchanged speed. (c) Using the affinity laws, find the speed, as a fraction of the original, that restores the original flow with the clogged filter, and the ratio of the air power $\Delta p\,\dot{V}$ at that point to the air power of the clean case.

Exercise 6.3. A hall's facility fans draw 15 kW at full speed. (a) Compute their power at 70 % speed. (b) A 30 kW rack receives 2.5 kg/s at full speed with a 22°C inlet; assuming its flow scales with fan speed and no recirculation, compute its rise and exhaust temperature at 70 % speed. (c) Compare the power of two identical fans at 50 % speed with one at full speed delivering the same total flow.

Exercise 6.4. A rack reports $P_{\text{rack}} = 24$ kW, $T_{\text{in}} = 25$°C and $T_{\text{out}} = 37$°C; the supply is at 19°C and the adjacent hot aisle at 37°C. (a) Estimate the rack airflow and the recirculation

fraction, and state which of the two numbers is measured, calculated or inferred. (b) With power measured to 2 % and each temperature to 0.5 K (one standard deviation), estimate the standard deviation of the airflow. (c) The exhaust sensor is later found to read 1.5 K above the mixed exhaust. Correct the airflow estimate, and state in which direction the original estimate was wrong.

6.9 Connection to the TERRA framework

The air side of a data hall is a conservation network in the plainest sense, and the chapter's model maps onto a graph directly.

- **Air volumes are mass and energy nodes.** The supply plenum, each cold aisle, each rack, each hot aisle and the return plenum is a node with two balances, equations (6.18) and (6.19); racks are the nodes where heat $\dot{Q}_i$ enters, and that heat is the edge flow arriving from the electrical network through the rack's power.
- **Paths are edges with constitutive laws.** Tiles, racks, grilles and leakage paths are edges carrying $\dot{m}_e = f(\Delta p_e, K_e)$, equation (6.17); fans are edge components with a curve, a speed bound and a power by the affinity laws. The network of Figure 6.6 is the graph itself.
- **Measured, calculated and inferred stay separate.** Inlet and exhaust temperatures, rack power and aisle pressure differences are measurements attached to nodes and edges; rack airflows, recirculation and bypass fractions and resistance coefficients are unmeasured parameters, and the framework returns them as posterior distributions, as in equation (6.16), rather than as point values.
- **Hot spots are ranked hypotheses.** The five explanations of rack 18 in Table 6.1 are alternative parameter shifts of the same graph (a load, a resistance, a recirculation fraction, a sensor bias), and the result of a diagnosis is a posterior ranking over them; the next measurement is chosen by its expected information, as in Table 6.2.
- **Risk is a probability per rack.** Local adequacy becomes $\Pr(\dot{Q}_{\mathrm{avail},i} < \dot{Q}_{\mathrm{rack},i})$, or equivalently $\Pr(T_{\mathrm{in},i} > T_{\max})$, computed from the posterior; the chance constraint of equation (6.21) is that probability used as a control limit.

In summary, a rack is cooled by the air that reaches it, at the temperature it arrives with; the delivery of that air is governed by a network of pressures, quadratic resistances and fans, degraded by recirculation and bypass, and described by node balances that a reduced model can solve quickly; and the airflow that decides each rack's margin is inferred, so the margin is a probability. The next chapter follows the heat out of the air and into the cooling equipment: coils, chillers, cooling towers and the water loops that connect them (Chapter 7).

Chapter 7

Cooling equipment: coils, pumps, chillers, towers and economizers

7.1 The heat-rejection chain

The previous chapters followed electrical power into the IT equipment and out again as heat in the room air (Chapter 5), and followed that air through the aisles to the cooling units (Chapter 6). The heat does not stop there. Every watt the servers consume must eventually leave the building, and the equipment that carries it from the room to the outdoors is the subject of this chapter. Its question is

once heat is in the room air or in a liquid loop, how does the facility move it outdoors, what limits how much it can move, and what can be measured and inferred along the way?

The chapter is a catalogue. It takes the equipment families one at a time, the cooling coil, the pumps, the chiller, the cooling tower and its alternatives, and describes each on the same template: what it does, the relation that governs it, what limits it, and what is measured on it. The last two body sections then put the families back together as one network with capacities, and ask how that network is measured, diagnosed, controlled and exposed to the weather.

Jargon: CRAC, CRAH, coil, evaporator, condenser, COP, temperature lift, wet-bulb temperature, economizer, fouling

A *CRAC* (computer room air conditioner) is a room cooling unit with its own refrigeration cycle, compressor included. A *CRAH* (computer room air handler) is a room cooling unit without a compressor: fans push room air across a coil fed with chilled water from a central plant. A *coil* is a heat exchanger made of finned tubes: water or refrigerant flows inside the tubes and air flows across the fins. The *evaporator* is the heat exchanger of a refrigeration cycle in which the refrigerant boils at low pressure and absorbs heat; the *condenser* is the one in which it condenses at high pressure and rejects heat. The *coefficient of performance* (COP) of a chiller is the heat it removes divided by the work it consumes. The *temperature lift* is the difference between the condensing and the evaporating temperatures of the refrigerant: how far the chiller must raise the heat in temperature. The *wet-bulb temperature* is the temperature a wetted surface reaches when air blows over it and evaporation cools it; it is lower than the ordinary (dry-bulb) temperature unless the air is saturated, and it is the lowest temperature evaporative cooling can approach. An *economizer* is an arrangement that uses cold outdoor air or water to cool the facility directly, reducing or eliminating compressor work. *Fouling* is the accumulation of dirt, scale, biological growth or deposits on a heat transfer surface, which adds thermal resistance and reduces the heat the surface can pass.

Key idea

Cooling equipment does not destroy heat; it moves it, through a chain of coupled heat-transport components from the rack to the atmosphere. Each link has a governing balance, a capacity, an efficiency, controls and a way to degrade, and the heat the facility can reject is set by the weakest link under the current weather, not by the nameplate of any single machine.

7.1.1 The reference chain from rack to atmosphere

The chapter uses one reference architecture, the most common one in large facilities:

$$\text{IT equipment} \to \text{room air} \to \text{CRAH} \to \text{chilled water} \to \text{chiller} \to \text{condenser water} \to \text{cooling tower} \to \text{atmosphere.}$$

Figure 7.1 draws it. The boxes are the places where heat changes carrier: the IT racks heat the room air; the CRAH coils pass that heat into chilled water; the chiller's evaporator passes it into a refrigerant; the chiller's condenser passes it, with the compressor's work added, into condenser water; the cooling tower passes it into the outdoor air. The arrows between boxes are the carriers. The upper arrow of each pair carries the warm stream toward the plant and the lower arrow carries the cooled stream back, so each pair is a closed loop: an air loop through the room, the chilled-water loop between the CRAHs and the chiller, the refrigerant circuit inside the chiller, and the condenser-water loop between the chiller and the tower. The temperatures on the arrows are the illustrative values used throughout the chapter. The labels $\dot{Q}_1$ to $\dot{Q}_5$ in the boxes name the heat flow that each stage passes on; Section 7.5.1 uses them to write the chain as a network.

Not every data center uses this architecture. Smaller rooms use CRAC units with their own compressors and outdoor condensers; some plants use air-cooled chillers and no condenser-water loop; some facilities cool with outdoor air for most of the year; liquid-cooled racks move heat from the chips into water without the room-air stage. The reference chain contains all the equipment families these variants are built from, so each variant is a subset or a rearrangement of it.

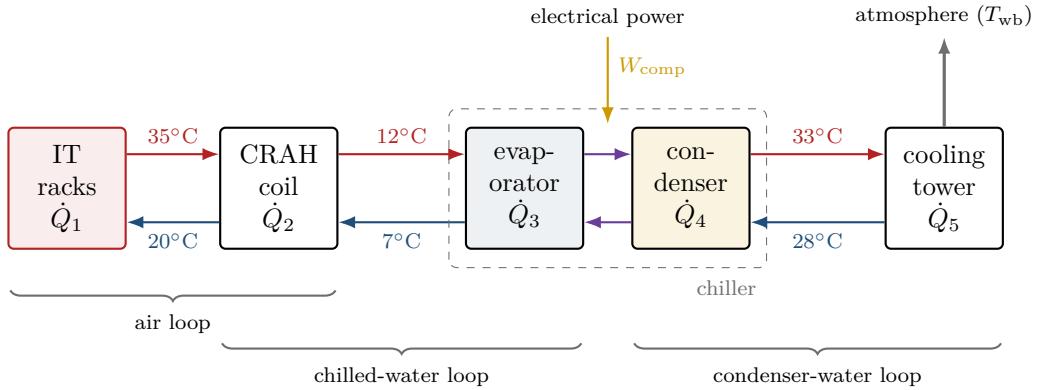


Figure 7.1: The reference heat-rejection chain. Each box is a place where heat changes carrier and each pair of arrows is a closed loop, warm stream above and cooled stream below. The chiller (dashed) contains the refrigerant circuit between its evaporator and condenser and is driven by compressor power W_{comp}. Temperatures are illustrative. $\dot{Q}_1$ to $\dot{Q}_5$ are the heat flows passed on by each stage.

7.1.2 CRAC and CRAH units

The two room cooling units of the jargon box differ in where the refrigeration happens. A CRAC contains a complete refrigeration cycle: a compressor, an evaporator coil over which the room air passes, an expansion device, and a condenser that is usually placed outdoors or cooled by a separate water or glycol loop. Warm room air enters the CRAC and leaves cool; the heat goes into the CRAC's refrigerant and is rejected by its own condenser. Each CRAC is a small self-contained plant, with its own compressor power and its own efficiency.

A CRAH contains no main compressor. Its fans move room air across a coil, and chilled water from a central plant flows through the coil's tubes; heat passes from the air to the water, and the water carries it away. The refrigeration machinery sits in the central chiller plant, shared by all CRAHs. The distinction matters for modeling and for measurement. In a CRAC facility the room heat and the compressor work are rejected unit by unit, and each unit's efficiency must be characterized separately. In a CRAH facility the room units are pure heat exchangers with fans, and the refrigeration, its efficiency and its sensitivity to the weather are concentrated in a few large chillers. The reference chain of Figure 7.1 is a CRAH architecture.

7.2 The coil and the chilled-water loop

This section describes the first heat exchanger of the chain, the CRAH coil, writes its balance from the air side and from the water side, introduces the overall conductance that limits it, and closes the chilled-water loop that connects the coils to the plant.

7.2.1 The cooling coil

A cooling coil is a heat exchanger with warm air on one side and cold water on the other. In the reference example, room air enters the coil at 35°C and leaves at 20°C, and chilled water enters the coil at $T_{\mathrm{CHW,s}} = 7°\mathrm{C}$ and leaves at $T_{\mathrm{CHW,r}} = 12°\mathrm{C}$. Heat flows from the air, through the fins and tube walls, into the water. In steady operation nothing accumulates in the coil, so the heat the air loses is, apart from small losses to the surroundings, the heat the water gains.

Figure 7.2 draws the coil as a counterflow exchanger. The upper part is a schematic: air flows from left to right through the upper channel and water flows from right to left through the lower channel, so the coldest water meets the coldest air at the right end and the warmest air meets the warmest water at the left end. The lower part plots temperature (vertical axis, °C) against position along the coil (horizontal axis, from the air inlet at 0 to the air outlet at 1). The air curve falls from 35 to 20°C; the water curve, read from right to left in the direction the water flows, rises from 7 to 12°C. The vertical gaps between the curves are the local temperature differences that drive the heat: 23 K at the air inlet end $(35 - 12)$ and 13 K at the air outlet end $(20 - 7)$. The air temperature changes three times as much as the water temperature because the water stream has three times the heat capacity rate, as the next subsection shows. Counterflow is the arrangement that lets the air leave colder than the water leaves: the outgoing air at 20°C is well above the incoming water, and the outgoing water at 12°C is well below the incoming air, so both streams use the full span between the two inlet temperatures.

7.2.2 The air-side and water-side balances

The heat the air gives up is its mass flow times its specific heat times its temperature drop,

$$\dot{Q}_{\mathrm{air}} = \dot{m}_{\mathrm{air}}\, c_{p,\mathrm{air}}\, (T_{\mathrm{air,in}} - T_{\mathrm{air,out}}). \tag{7.1}$$

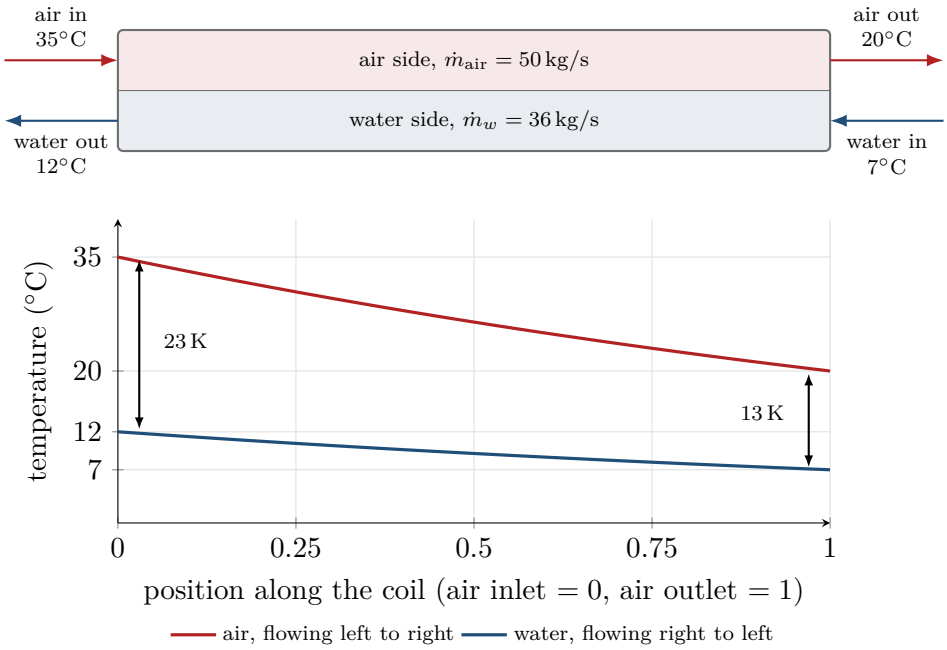


Figure 7.2: The CRAH coil as a counterflow heat exchanger with its four temperatures. Top: schematic of the two streams. Bottom: temperature profiles along the coil for the illustrative example. The local difference $\theta(x) = 23\,(13/23)^x$ decays exponentially from 23 to 13 K; the air temperature is $35 - 15f(x)$ and the water temperature $12 - 5f(x)$ with $f(x) = (23 - \theta(x))/10$ the fraction of the heat transferred up to position x.

With an illustrative air flow $\dot{m}_{\text{air}} = 50\,\text{kg/s}$ (about $42\,\text{m}^3/\text{s}$ at a density of $1.2\,\text{kg/m}^3$), $c_{p,\text{air}} \approx 1.0\,\text{kJ/(kg K)}$ and a temperature drop of $35 - 20 = 15\,\text{K}$,

$$\dot{Q}_{\text{air}} = 50\,\text{kg/s} \times 1.0\,\text{kJ/(kg K)} \times 15\,\text{K} = 750\,\text{kW}.$$

The heat the water gains has the same form, with the water's temperature rise,

$$\dot{Q}_{\text{water}} = \dot{m}_w\, c_{p,w}\, (T_{\text{CHW,r}} - T_{\text{CHW,s}}). \tag{7.2}$$

With $\dot{m}_w = 36\,\text{kg/s}$, $c_{p,w} = 4.18\,\text{kJ/(kg K)}$ and a rise of $12 - 7 = 5\,\text{K}$,

$$\dot{Q}_{\text{water}} = 36 \times 4.18 \times 5 = 752.4 \approx 752\,\text{kW}.$$

The products $\dot{m}_{\text{air}} c_{p,\text{air}} = 50\,\text{kW/K}$ and $\dot{m}_w c_{p,w} = 150.5\,\text{kW/K}$ are the *heat capacity rates* of the two streams: the heat each absorbs per kelvin of temperature change. Because the water's is about three times the air's, the same heat changes the water temperature by about a third as much, 5 K against 15 K, which is what Figure 7.2 shows. Water is the better carrier: its specific heat is about four times that of air and its density about 800 times, so a pipe carries heat that would need a large duct of air.

The two estimates, 750 and 752 kW, differ by 2 kW, about 0.3 %, which is well inside the uncertainty of any field measurement of the quantities involved. This agreement is not a coincidence of the example but a consequence of energy conservation, and it turns into a test. $\dot{Q}_{\text{air}}$ and $\dot{Q}_{\text{water}}$ are both *calculated* quantities, built from different *measured* quantities: air temperatures and an air flow on one side, water temperatures and a water flow on the other. If the two calculations disagree by more than their combined uncertainty, then at least one of the measurements, or the assumption that nothing else exchanges heat with the coil, is wrong. Possible causes include a biased temperature sensor, an air-flow estimate inferred from fan

speed that no longer holds, a water-flow meter out of calibration, moisture condensing on the coil (which adds latent heat to the air side that equation (7.1) does not count), or air bypassing the coil.

Principle 7.1. *The air side and the water side of a coil are two independent calculations of the same heat. Agreement within measurement uncertainty is the expected state; disagreement beyond it is evidence of a sensor fault, an unmodeled heat path or a wrong assumption, and deserves investigation.*

7.2.3 Overall conductance and fouling

The balances of equations (7.1) and (7.2) say how much heat the coil passed, given the temperatures. They do not say how much heat the coil *can* pass. That is set by a third relation, the heat-exchanger equation,

$$\dot{Q} = UA\,\Delta T_{\mathrm{lm}}, \tag{7.3}$$

where U is the overall heat-transfer coefficient (watts per square meter per kelvin), A the heat-transfer area, and their product UA the overall conductance of the exchanger in watts per kelvin. The driving temperature difference is the log-mean temperature difference,

$$\Delta T_{\mathrm{lm}} = \frac{\Delta T_1 - \Delta T_2}{\ln(\Delta T_1/\Delta T_2)}, \tag{7.4}$$

where ΔT_1 and ΔT_2 are the stream-to-stream differences at the two ends. It is the correct average of the local difference because that difference decays exponentially along a counterflow coil, as in Figure 7.2. Equation (7.3) reads as

$$\text{heat transferred} \propto \text{heat-transfer capability} \times \text{temperature driving force}.$$

For the example, $\Delta T_1 = 23\,\mathrm{K}$ and $\Delta T_2 = 13\,\mathrm{K}$, so

$$\Delta T_{\mathrm{lm}} = \frac{23-13}{\ln(23/13)} = \frac{10}{0.571} = 17.5\,\mathrm{K}, \qquad UA = \frac{750\,\mathrm{kW}}{17.5\,\mathrm{K}} = 42.8\,\mathrm{kW/K}.$$

UA is not measured by any sensor. It is *inferred* from measured temperatures and a measured or calculated heat, and it is a property of the coil's physical condition.

That condition changes. Dust on the air-side fins, scale or biological film on the water-side tube walls, and corrosion all add thermal resistance in series with the clean coil and lower U. At the same temperatures, a lower UA passes proportionally less heat. In operation the temperatures do not stay the same: with the same inlet temperatures and flows, a fouled coil cools the air less, which widens the temperature differences and partly compensates. For the example coil, a 20 % loss of conductance, from 42.8 to 34.2 kW/K, computed with the standard effectiveness relation for counterflow exchangers at the same air and water inlet temperatures and flows, lowers the heat from 750 to about 651 kW, a 13 % loss, and raises the air leaving the coil from 20 to about 22°C; the water leaves at about 11.3°C instead of 12°C. Nothing in the fan speed or the water flow reveals the change. A fouled coil degrades cooling while the air and water flows look normal, and it shows up only as a slow drift in the temperatures, which is why UA is worth tracking as an inferred state.

7.2.4 The chilled-water loop

The coils are the loads of a closed water circuit, the chilled-water loop. The chiller produces cold supply water at $T_{\mathrm{CHW,s}}$; pumps distribute it through pipes to the CRAH coils; each coil warms it; the warm return water at $T_{\mathrm{CHW,r}}$ flows back to the chiller, which cools it again. Water

is the carrier: it picks up heat in the data hall and delivers it to the plant, which may be in a separate building. The loop holds a large mass of water, several hundred cubic meters in a large facility, and that mass is a thermal store that buffers short mismatches between the heat entering at the coils and the heat removed at the chiller.

Applied to the whole loop, equation (7.2) links the three quantities that operators watch on the plant screen: the total heat delivered to the chiller equals the total water flow times $c_{p,w}$ times the loop temperature difference $T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}}$. For a given heat, a larger flow gives a smaller difference and a smaller flow a larger one. A loop designed for a 5 K difference that runs at 3 K is carrying the same heat with two thirds more water than intended, which costs pump energy (Section 7.3.2) and lowers the chiller's efficiency; this condition is common enough to have a name in the industry, the low-ΔT problem. The loop difference is therefore both a performance indicator and a diagnostic quantity, and Section 7.6.3 uses a rise in the return temperature as its worked example.

7.3 Pumps and the chiller

This section covers the two machines that do work on the chilled-water loop: the pumps, which move the water, and the chiller, which lifts the heat from the loop to a warmer sink. For each it gives the governing relation, the energy it consumes and what limits it.

7.3.1 System and pump curves

Water in the loop must be pushed through pipes, bends, valves, coils, the chiller's evaporator, strainers and changes of elevation. Each of these resists the flow, and the pressure difference needed to drive a volumetric flow $\dot{V}$ through the whole circuit grows with the square of the flow, as for the air paths of Chapter 6:

$$\Delta p_{\mathrm{system}} \approx K\,\dot{V}^2, \tag{7.5}$$

with K a flow-resistance coefficient that lumps every component of the circuit. In an open circuit lifting water to a height, a static term is added; in a closed loop the elevations cancel and the quadratic term dominates. A pump supplies a pressure difference that falls as its flow rises, its *pump curve*. The loop runs where the two meet,

$$\Delta p_{\mathrm{pump}}(\dot{V}) = \Delta p_{\mathrm{system}}(\dot{V}), \tag{7.6}$$

exactly as a fan and its duct system settle at the intersection of the fan curve and the system curve.

As an illustrative example, take a chilled-water loop with $K = 4340\,\mathrm{kPa\,s^2/m^6}$ and a pump whose curve is $\Delta p_{\mathrm{pump}} = 400 - 2600\,\dot{V}^2$ (kPa, with $\dot{V}$ in $\mathrm{m^3/s}$). Setting the two equal gives $\dot{V}^2 = 400/(4340 + 2600) = 0.0576$, so $\dot{V} = 0.24\,\mathrm{m^3/s}$ at $\Delta p = 4340 \times 0.0576 = 250\,\mathrm{kPa}$. At $1000\,\mathrm{kg/m^3}$ that is $240\,\mathrm{kg/s}$ of water, and by equation (7.2) it carries $240 \times 4.18 \times 5 \approx 5.0\,\mathrm{MW}$ at a 5 K loop difference: the flow a 5 MW data hall needs. Closing a valve raises K and moves the operating point to a lower flow; a clogged strainer does the same, unintentionally.

7.3.2 Pump power and the affinity laws

The power delivered to the water is the pressure difference times the volumetric flow, $P_{\mathrm{hyd}} = \Delta p\,\dot{V}$, and the electrical power the pump draws is larger by the inverse of its efficiency η, which combines the hydraulic, mechanical, motor and drive losses:

$$P_{\mathrm{pump}} = \frac{\Delta p\,\dot{V}}{\eta}. \tag{7.7}$$

At the operating point of the example, with an illustrative overall efficiency $\eta = 0.75$,

$$P_{\text{hyd}} = 250\,000\,\text{Pa} \times 0.24\,\text{m}^3/\text{s} = 60\,\text{kW}, \qquad P_{\text{pump}} = 60/0.75 = 80\,\text{kW}.$$

The trade-off is direct: more water flow carries more heat at a given temperature difference, and costs more pump energy.

Affinity laws. For a given pump and a system curve through the origin, flow, pressure and power scale with the rotational speed N_{pump} as

$$\dot{V} \propto N_{\text{pump}}, \qquad \Delta p \propto N_{\text{pump}}^2, \qquad P_{\text{pump}} \propto N_{\text{pump}}^3. \tag{7.8}$$

The cube is the reason variable-speed pumping saves energy. Running the example pump at 80 % speed gives $\dot{V} = 0.8 \times 0.24 = 0.192\,\text{m}^3/\text{s}$, $\Delta p = 0.64 \times 250 = 160\,\text{kPa}$ and $P_{\text{pump}} = 0.512 \times 80 = 41\,\text{kW}$: a fifth less flow for about half the power. Two cautions apply. A real loop is rarely a pure quadratic through the origin, because controls hold a minimum pressure difference across the farthest coil, so the savings at low speed are smaller than the cube predicts. And flow is what carries the heat: at $0.192\,\text{m}^3/\text{s}$ the same 5 MW needs a loop difference of $5000/(192 \times 4.18) = 6.2\,\text{K}$ rather than 5 K. Whether the coils can deliver that larger difference depends on their *UA* and the air temperatures; if they cannot, the heat transfer is insufficient and the room warms. Too little flow saves pump energy at the expense of cooling.

7.3.3 The chiller and the refrigeration cycle

The chiller receives the return water at 12°C and sends supply water back at 7°C. To do that it must move heat from the chilled-water loop to a sink, the condenser water or the outdoor air, that is warmer than the chilled water. Heat flows spontaneously only from warm to cold, so moving it the other way takes work, and the chiller is a machine that spends electrical work to move heat uphill in temperature. It moves heat; it does not destroy it.

It does so with a vapor-compression refrigeration cycle of four components. Figure 7.3 draws the cycle as a loop of refrigerant passing through them in order. The box at the bottom is the *evaporator*: liquid refrigerant at low pressure and low temperature boils inside it, absorbing heat $\dot{Q}_{\text{evap}}$ from the chilled water, which leaves colder. The box on the right is the *compressor*: it takes the low-pressure vapor and compresses it to high pressure, consuming electrical power W_{comp} and raising the vapor's temperature. The box at the top is the *condenser*: the hot high-pressure vapor condenses back to liquid, rejecting heat $\dot{Q}_{\text{cond}}$ to the condenser water. The box on the left is the *expansion device*, a valve or orifice that drops the liquid back to low pressure, which cools it below the chilled-water temperature and closes the loop. The three thick external arrows are the only energy exchanges with the outside: heat in at the evaporator, work in at the compressor, heat out at the condenser.

7.3.4 The energy balance of a chiller

The chiller is a control volume with two energy inputs and one output. In steady operation the stored energy does not change, so what leaves equals what enters:

$$\dot{Q}_{\text{cond}} = \dot{Q}_{\text{evap}} + W_{\text{comp}}. \tag{7.9}$$

With $\dot{Q}_{\text{evap}} = 5\,\text{MW}$ and $W_{\text{comp}} = 1\,\text{MW}$,

$$\dot{Q}_{\text{cond}} = 5\,\text{MW} + 1\,\text{MW} = 6\,\text{MW}.$$

The consequence for plant design is that the heat-rejection equipment downstream of the chiller must handle more than the IT heat. In the example, a data hall that produces 5 MW

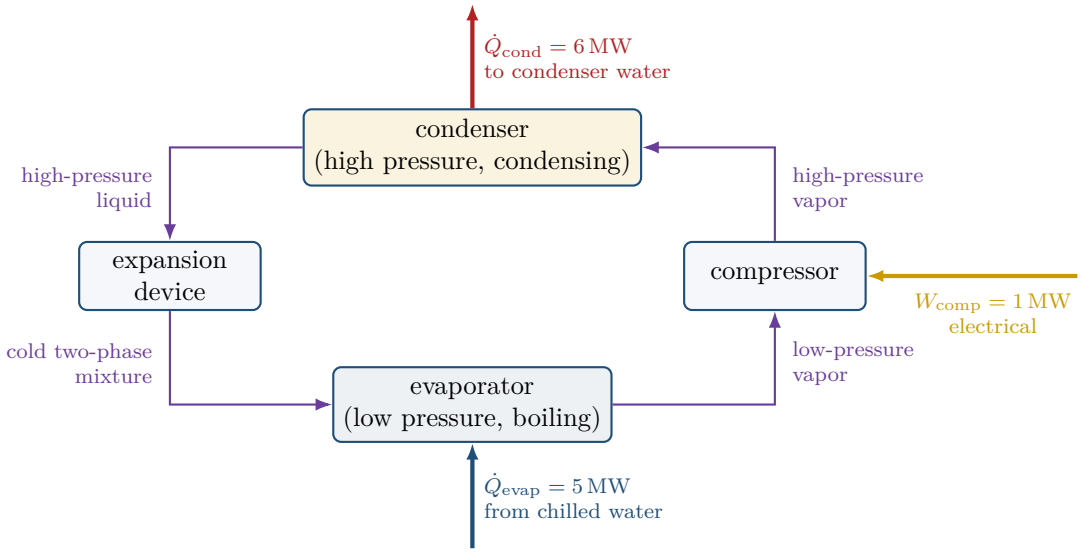


Figure 7.3: The vapor-compression refrigeration cycle of a chiller, with the illustrative values of Section 7.3.4. The refrigerant (purple) circulates through four components; heat $\dot{Q}_{\text{evap}}$ enters at the evaporator, work W_{comp} enters at the compressor, and their sum $\dot{Q}_{\text{cond}} = \dot{Q}_{\text{evap}} + W_{\text{comp}}$ leaves at the condenser.

of heat requires towers, condenser-water pumps and condenser-water piping for 6 MW. The compressor's electrical power does not disappear; it arrives at the condenser as heat, together with the heat it moved. Equation (7.9) is also a measurable balance: the evaporator heat is calculated from the chilled-water flow and its temperature difference, the condenser heat from the condenser-water flow and its difference, and the compressor power is metered; the three should close within their uncertainties.

7.3.5 Coefficient of performance

The figure of merit of a chiller is its coefficient of performance, the cooling delivered per unit of work consumed,

$$\text{COP} = \frac{\dot{Q}_{\text{evap}}}{W_{\text{comp}}}. \tag{7.10}$$

For the example, $\text{COP} = 5\,\text{MW}/1\,\text{MW} = 5$: each megawatt of compressor power removes five megawatts of heat from the chilled water. Higher is better. A chiller plant cooling 5 MW of IT heat at $\text{COP} = 5$ spends 1 MW on compressors; at $\text{COP} = 4$ it would spend 1.25 MW, and at $\text{COP} = 6$, 0.83 MW.

The COP is not an efficiency. A value of 5 does not violate energy conservation, because a COP is not a ratio of energy out to energy in of the same conversion. An efficiency compares the useful output of a conversion with its input, and cannot exceed one. A chiller does not convert 1 MW of electricity into 5 MW of cooling; it uses 1 MW of work to *move* 5 MW of heat that already exists, from a cold stream to a warm one, and delivers $5 + 1 = 6$ MW at the condenser. Every term is accounted for by equation (7.9). The same reasoning applies to a heat pump, whose COP for heating is $\dot{Q}_{\text{cond}}/W_{\text{comp}} = 6$ in the example. The COP compares a transported quantity with the work needed to transport it, and a ratio above one is the normal state of any refrigeration machine.

> **Principle 7.2.** *A chiller moves heat; it does not create or destroy it. The heat it rejects is the heat it removes plus the work it consumes, and its COP, which is commonly between three and seven, is a transport ratio, not an efficiency.*

7.3.6 Temperature lift

How much work it takes to move a unit of heat depends on how far it is moved in temperature. The *temperature lift* is

$$\text{lift} = T_{\text{cond}} - T_{\text{evap}}, \tag{7.11}$$

the difference between the refrigerant's condensing temperature, which sits a few kelvin above the condenser water, and its evaporating temperature, which sits a few kelvin below the chilled water. The thermodynamic upper bound for the COP of any machine moving heat between those two temperatures is the Carnot value $T_{\text{evap}}/(T_{\text{cond}} - T_{\text{evap}})$, with temperatures in kelvin, and real chillers reach a roughly constant fraction of it over their normal range. An illustrative model is

$$\text{COP} \approx \eta_{\text{C}} \, \frac{T_{\text{evap}}}{T_{\text{cond}} - T_{\text{evap}}}, \tag{7.12}$$

with η_{C} the fraction of the Carnot value achieved. The lift is in the denominator: a larger lift means more compressor work for the same heat and a lower COP.

Figure 7.4 plots equation (7.12). The horizontal axis is the lift in kelvin; the vertical axis is the COP. The curve uses $\eta_{\text{C}} = 0.5$ and an evaporating temperature of 278 K (5°C). The middle point is the reference chiller of Figure 7.3, COP = 5 at a lift of 27.8 K, for example evaporating at 5°C and condensing at about 33°C. The left point is the same chiller with the lift reduced by 3 K: COP 5.6, so the compressor power for 5 MW falls from 1.00 to 0.89 MW. The right point is a hot afternoon that raises the condensing temperature by 8 K: lift 35.8 K, COP 3.9, compressor power 1.29 MW and condenser heat 6.29 MW. The curve is steep at small lifts and flat at large ones, so each kelvin of lift saved is worth most when the lift is already small.

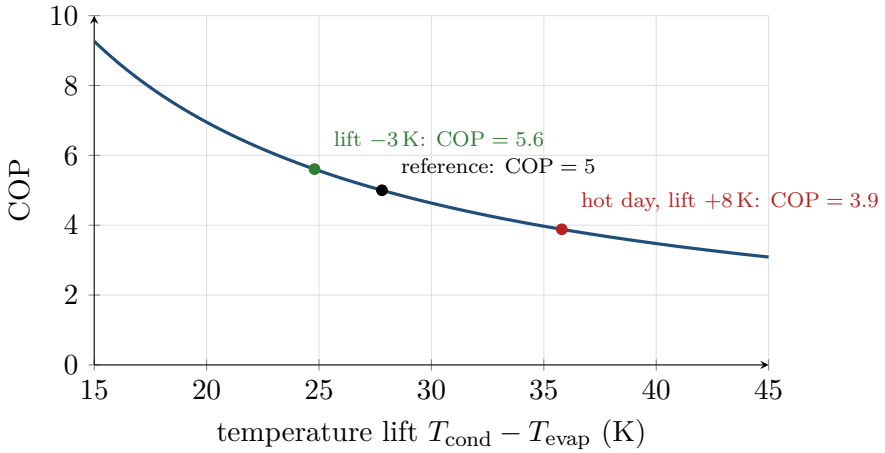


Figure 7.4: Coefficient of performance against temperature lift, from the illustrative closed form $\text{COP} = \eta_{\text{C}} \, T_{\text{evap}}/(T_{\text{cond}} - T_{\text{evap}})$ with $\eta_{\text{C}} = 0.5$ and $T_{\text{evap}} = 278$ K held fixed. Reducing the lift raises the COP; a hot day raises the lift and lowers it.

Two operating conditions raise the lift. Colder chilled water lowers the evaporating temperature, and hotter outdoor conditions raise the condensing temperature. Both raise W_{comp} and lower the COP. The operator controls the first and not the second. Raising the chilled-water supply temperature, for example from 7 to 10°C, raises the evaporating temperature by about 3 K and gives the left-hand point of Figure 7.4; because the evaporating temperature also appears in the numerator of equation (7.12), the full expression gives slightly more, $0.5 \times 281/24.8 = 5.7$. This is why modern facilities run warmer chilled water than older ones. The limit is the IT equipment: warmer chilled water means warmer supply air from the CRAH coils, which is acceptable only while the server inlet temperatures stay within their allowed range, and it also

lowers the coil's driving temperature difference, so the coils need more water or more area for the same heat.

7.4 Heat rejection to the atmosphere

The chiller delivers the heat, now increased by the compressor work, to the condenser. This section follows it to the outdoors: through the condenser-water loop and the cooling tower, through the physics of evaporation and the wet-bulb temperature that limits it, and through the alternatives that reject heat without a tower or without a compressor.

7.4.1 The condenser-water loop

In a water-cooled plant the condenser is cooled by a second water circuit, the condenser-water loop. Refrigerant condensing in the chiller heats the condenser water; condenser-water pumps send it to the cooling towers; the towers cool it; it returns to the condenser. In the reference example of Figure 7.1 the condenser water leaves the chiller at about 33°C and returns at 28°C. With $\dot{Q}_{\text{cond}} = 6\,\text{MW}$ and a 5 K difference, equation (7.2) gives the condenser-water flow,

$$\dot{m}_{\text{CW}} = \frac{6000\,\text{kW}}{4.18\,\text{kJ/(kg K)} \times 5\,\text{K}} = 287\,\text{kg/s},$$

a fifth more than the chilled-water flow, because it carries the compressor heat as well.

The plant therefore has two water loops, and they must not be confused. The *chilled-water loop* runs between the CRAH coils and the chiller's evaporator, at temperatures around 7 to 12°C (or warmer in modern designs), and carries the IT heat. The *condenser-water loop* runs between the chiller's condenser and the cooling towers, at temperatures around 28 to 33°C, and carries the IT heat plus the compressor work. The two loops never mix: the refrigerant inside the chiller passes the heat from one to the other. A temperature reading labeled only "supply" or "return" is ambiguous until it is assigned to one loop, and an error in that assignment is a topology error of the kind Section 7.6.4 discusses.

7.4.2 The cooling tower and latent heat

A cooling tower brings warm condenser water into direct contact with outdoor air. Figure 7.5 draws a typical induced-draft tower in cross-section. Warm condenser water enters at the top and is sprayed over a *fill*, a packing of plastic sheets or splash bars that spreads it into thin films and droplets with a large surface. A fan at the top draws outdoor air in through openings near the bottom and up through the fill, against the falling water. A small fraction of the water evaporates into the air, and the evaporation takes heat from the rest of the water, which falls into the basin cooler and returns to the condenser. The air leaves at the top warmer and nearly saturated with water vapor. Makeup water replaces what evaporates and what is drained to limit the concentration of dissolved minerals.

The physics that makes a tower effective is the difference between two ways of carrying heat. *Sensible* heat changes the temperature of a substance without changing its phase,

$$\dot{Q}_{\text{sensible}} = \dot{m}\, c_p\, \Delta T, \tag{7.13}$$

as in every balance so far. *Latent* heat changes its phase at constant temperature,

$$\dot{Q}_{\text{latent}} = \dot{m}_{\text{evap}}\, L_v, \tag{7.14}$$

where $\dot{m}_{\text{evap}}$ is the rate at which water evaporates and L_v is the latent heat of vaporization of water, about 2400 kJ/kg at tower temperatures. One kilogram of water cooled by 5 K gives up

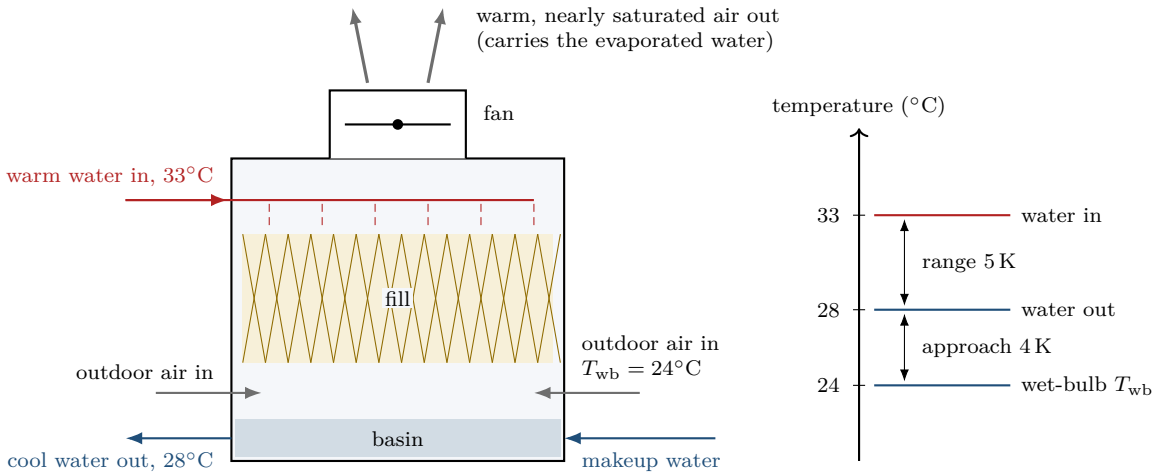


Figure 7.5: An induced-draft cooling tower in cross-section (left) and its temperatures (right), illustrative. Warm condenser water is sprayed over the fill and falls against rising outdoor air; evaporation of a small fraction cools the rest. The *range* is the water's temperature drop; the *approach* is how far the cooled water stays above the outdoor wet-bulb temperature.

$4.18 \times 5 = 21\,\mathrm{kJ}$; one kilogram evaporated carries away 2400 kJ, more than a hundred times as much. A small evaporated mass therefore carries a large heat. For the 6 MW of the example, if the whole heat left as latent heat,

$$\dot{m}_{\mathrm{evap}} = \frac{6000\,\mathrm{kW}}{2400\,\mathrm{kJ/kg}} = 2.5\,\mathrm{kg/s},$$

about 0.9 % of the 287 kg/s condenser-water flow, or roughly $9\,\mathrm{m}^3$ of water per hour. In practice most of the tower's heat leaves this way and a smaller part leaves as sensible heating of the air. The water consumption of a water-cooled data center is mostly this evaporation, plus the drain water.

7.4.3 Humidity and the wet-bulb temperature

Evaporation depends on how much more water vapor the air can take up. Dry air accepts vapor readily and permits strong evaporation; humid air is closer to saturation and permits less; saturated air permits none. The wet-bulb temperature T_{wb} of the jargon box combines the air's temperature and humidity into the one number that matters for evaporation: it is the temperature at which water evaporating into the air is in balance with the heat the air supplies to it, and it is the lowest temperature to which evaporation can cool water in that air. Tower performance therefore follows the wet-bulb temperature much more closely than the ordinary dry-bulb temperature T_{amb}.

The two can differ widely. Air at 35°C and 40 % relative humidity has a wet-bulb temperature of about 24°C; air at the same 35°C and 70 % relative humidity has a wet-bulb temperature of about 30°C. A tower can cool its water only toward the wet-bulb temperature, never to it, and the remaining gap is the *approach* of Figure 7.5. With an approach of 4 K, the dry afternoon gives condenser water at $24 + 4 = 28$°C and the humid one at $30 + 4 = 34$°C, six kelvin warmer at the same dry-bulb temperature. A thermometer in the shade does not tell how well the towers will work; a humidity sensor beside it does.

7.4.4 Outdoor conditions and available capacity

The chain from the weather to the plant is now complete. On a cool day, at 15°C outdoors, heat rejection is easy: the towers produce cold condenser water, the condensing temperature is low, the lift is small, and the chillers run at high COP with capacity to spare. On a hot and humid day, at 38°C with high humidity, the wet-bulb temperature is high, the towers return warmer condenser water, the condensing temperature rises, the lift rises, the COP falls, and the compressors must work harder for the same heat. At the extreme, the compressors reach their maximum power or their pressure limits before the required heat is moved, and the available cooling falls. The same facility, with the same installed equipment, has different usable cooling capacity on different days.

A capacity relationship. The available cooling capacity is therefore a function, not a number,

$$\dot{Q}_{\text{cool,max}} = f(T_{\text{amb}},\ \text{RH},\ \text{equipment availability},\ \dot{m}_w,\ \text{equipment state}), \tag{7.15}$$

where equipment availability says which chillers, pumps and towers are in service, $\dot{m}_w$ is the water flow the pumps can provide, and the equipment state includes fouling and wear. The IT load the facility can carry is bounded by it, $P_{\text{IT}}^{\max} \lesssim \dot{Q}_{\text{cool,max}}$, because nearly all IT power becomes heat that must be removed (the small difference is the heat leaving through the building envelope and the other loads sharing the plant). Since the weather enters equation (7.15), a weather forecast becomes a forecast of cooling capacity, and that becomes a forecast of how much IT load can be carried safely:

$$\text{weather forecast} \rightarrow \text{cooling-capacity forecast} \rightarrow \text{IT capacity}.$$

This is a forecasting-and-risk problem rather than a design calculation, and Section 7.6.6 returns to it.

7.4.5 Air-cooled chillers and dry coolers

Not every plant has towers. An *air-cooled chiller* rejects its condenser heat directly to outdoor air: the condenser is a finned coil with fans blowing outdoor air across it, and there is no condenser-water loop at all. The plant then needs no condenser-water pumps and no cooling towers, uses far less water, and is simpler to maintain. The price is performance in hot weather. Without evaporation, the condenser can approach only the dry-bulb temperature, which is higher than the wet-bulb, and air is a poorer heat carrier than water, so the condensing temperature sits further above the outdoor temperature. On a 35°C afternoon an air-cooled chiller may condense at 50°C or more, while a water-cooled one condenses at 35 to 38°C; by Figure 7.4, that difference in lift is a large difference in COP.

Dry coolers. A *dry cooler* is a fluid-to-air heat exchanger: water or a water-glycol mixture flows through a finned coil and outdoor air blown across it cools the fluid, without evaporation and without a compressor of its own. It uses little or no water, and it can serve as the heat sink of a warm-water loop or as an economizer. Its rejection depends entirely on the dry-bulb temperature, through equation (7.3), and its driving difference shrinks as the outdoor temperature approaches the fluid temperature. For an illustrative dry cooler with fluid at 30°C, the driving difference is about 5 K at 25°C outdoors and about 2 K at 28°C; with the same UA the rejected heat falls by about 60 %, and above 30°C the dry cooler cannot cool the fluid at all.

7.4.6 Economizers

An economizer reduces or eliminates compressor work when outdoor conditions are favorable. Its operation is often called "free cooling", which is not exact: fans and pumps still consume energy, and only the compressor is relieved. There are two families.

- An *air-side economizer* brings cool outdoor air into the facility, filtered and sometimes humidified or dehumidified, in place of or mixed with recirculated air, and exhausts the warm air. The room air loop is then partly an open path to the outdoors.
- A *water-side economizer* uses cold outdoor conditions to cool the chilled-water loop through a heat exchanger, usually with water from the cooling towers or from a dry cooler, bypassing the chiller entirely or precooling the return water so that the chiller does less of the work.

In both cases W_{comp} falls, to zero in full economizer mode.

As an illustrative example, take a facility whose chilled-water supply requirement is 18°C, a warm-water design, on a day at 5°C outdoors. The outdoor conditions are 13 K below the requirement. The towers can produce water within a few kelvin of the wet-bulb temperature, which on such a day is a few degrees below 5°C, and a plate heat exchanger with an approach of about 2 K passes that cold to the chilled-water loop, so water well below 18°C is available with the chillers off. The chiller power falls from about 1 MW (for 5 MW of IT heat at COP = 5) to zero, leaving only the tower fans and pumps. If such conditions hold for an illustrative 4000 hours a year, the saving is about 4000 MWh of compressor energy. In suitable climates the savings are substantial, and they grow with the chilled-water temperature the IT equipment tolerates, because a warmer requirement makes more hours of the year cold enough. Between full economizer and full mechanical cooling there is a partial mode, in which the economizer precools the return water and the chiller removes the rest.

Comparison of the options. Table 7.1 compares the four ways of rejecting heat described in this section. The columns give the heat sink and the outdoor temperature that limits it, whether a compressor runs, the water consumed, and the weakness in hot or humid weather. The water-cooled plant with towers has the lowest lift and is limited by the wet-bulb temperature; the air-cooled options save water and are limited by the dry-bulb temperature; the economizer removes the compressor but works only when the outdoors is colder than the required supply temperature.

7.5 The plant as a network with capacities

The equipment families have been described one at a time. This section puts them back together as a network of heat flows, shows that the capacity of the network is set by its weakest link, and replaces the single nameplate capacity by a probability that the network can carry the load.

7.5.1 The chain as a network of heat flows

The reference chain of Figure 7.1 is a series of heat flows between carriers:

$$\text{rack} \xrightarrow{\dot{Q}_1} \text{room air} \xrightarrow{\dot{Q}_2} \text{chilled water} \xrightarrow{\dot{Q}_3} \text{refrigerant}$$
$$\xrightarrow{\dot{Q}_4} \text{condenser water} \xrightarrow{\dot{Q}_5} \text{atmosphere.}$$

Each carrier is a node with a conservation law, and each heat flow is an edge between two nodes. In steady state no node accumulates heat, so the flows that pass through without added work are equal,

$$\dot{Q}_1 \approx \dot{Q}_2 \approx \dot{Q}_3, \qquad \dot{Q}_4 = \dot{Q}_3 + W_{\text{comp}}, \qquad \dot{Q}_5 \approx \dot{Q}_4. \tag{7.16}$$

Table 7.1: Heat-rejection options compared.

Option	Heat sink and limiting temperature	Compressor	Water use	Weakness in hot or humid weather
Water-cooled chiller with cooling tower	outdoor air by evaporation; wet-bulb T_{wb}	yes	high: evaporation, about 1 % of the condenser flow, plus drain	high T_{wb} warms the condenser water: lift up, COP down, capacity down
Air-cooled chiller	outdoor air, sensible only; dry-bulb T_{amb}	yes	none for heat rejection	high T_{amb} raises the condensing temperature further than a tower would; lower COP, derating
Dry cooler	outdoor air through a finned coil, no evaporation; dry-bulb T_{amb}	none of its own	very low	driving difference shrinks as T_{amb} approaches the fluid temperature; no capacity above it
Economizer (air-side or water-side)	cold outdoor air directly, or through a heat exchanger; T_{amb} or T_{wb}	reduced or off	air-side none; water-side as the tower	works only when outdoors is colder than the required supply temperature; air-side also admits humidity and particles

The approximate signs absorb the small heat gains and losses along the way: the CRAH fan power adds to the room air, the chilled-water pumps add their power to the water, and pipes exchange heat with their surroundings. The one exact jump is at the chiller, where the compressor work enters, equation (7.9). During transients the equalities fail: a room warming or a water loop cooling stores heat, and the difference between inflow and outflow at a node is the rate of change of its stored energy, which is the subject of Chapter 8.

Written this way, the cooling plant has the same structure as any network of conserved flows. The nodes are the room air, the two water loops and the refrigerant; the edges are the coils, the evaporator, the condenser and the towers; each edge has a law that relates its flow to the states of the nodes it joins, such as equation (7.3) for a heat exchanger, and a limit on how much it can carry.

7.5.2 Component capacities and the bottleneck

Every component in the chain has a maximum heat it can pass. As an illustrative example, take a plant whose CRAH coils can together carry 10 MW, whose chillers can carry 12 MW and whose cooling towers can carry 9 MW, all expressed as heat carried for the IT load. Heat passes through the components in series, so the chain can carry no more than its smallest link:

$$\dot{Q}_{\mathrm{chain,max}} = \min(\dot{Q}_{\mathrm{CRAH,max}},\ \dot{Q}_{\mathrm{chiller,max}},\ \dot{Q}_{\mathrm{tower,max}}) = \min(10, 12, 9) = 9\,\mathrm{MW}. \qquad (7.17)$$

Figure 7.6 draws the three capacities as bars. The horizontal axis lists the three component groups in the order the heat passes through them; the vertical axis is the capacity in megawatts. The dashed line at 9 MW is the capacity of the chain. The chiller bar rises 3 MW above it, but that installed chiller capacity cannot be used: the towers cannot reject the heat. The towers are the bottleneck. A statement that the plant has 12 MW of chillers is true and does not answer the question of how much IT load the plant can cool.

The basis of each rating matters. If the tower's 9 MW is a rating for heat *rejected* rather than for IT heat carried, the towers must also reject the compressor work, by equation (7.9). At COP = 5 each megawatt of evaporator heat brings 0.2 MW of compressor heat, so the towers support only $\dot{Q}_{\text{evap}} = 9/(1+1/5) = 7.5$ MW, and the bottleneck is tighter than it first appeared. A capacity comparison is meaningful only once every rating is expressed on the same basis.

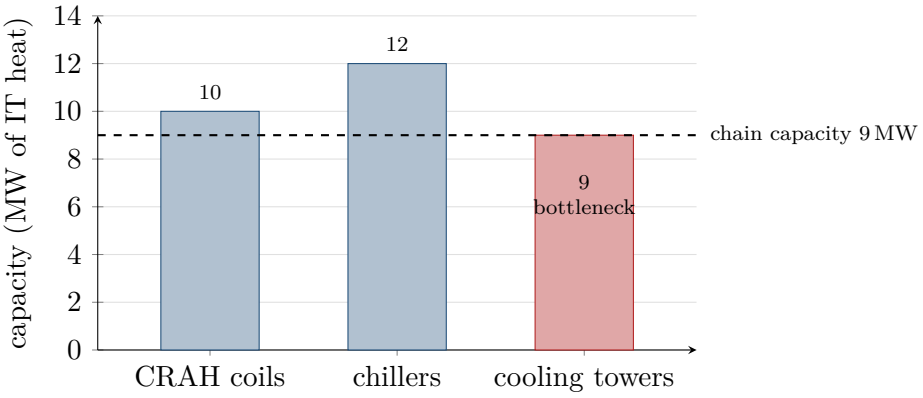


Figure 7.6: Illustrative capacities of the three component groups of the chain, in megawatts of IT heat. Heat passes through the groups in series, so the chain carries at most the smallest capacity, $\min(10, 12, 9) = 9$ MW (dashed); the 3 MW of chiller capacity above that line cannot be used.

Analogy with transmission capacity. The situation has an exact counterpart in a power system: plenty of generation is of no use to a load center if the transmission lines between them cannot carry the power. The cooling plant can likewise have ample nominal cooling production, chillers, while the heat cannot be delivered because of insufficient pump capacity, insufficient coil capacity, a pipe that limits the water flow, tower capacity, or local airflow that does not bring the room heat to the coils (Chapter 6). In both cases the deliverable quantity is a property of the paths as well as the sources. The local limits matter as much as the plant totals: a hall can have enough chilled water in aggregate and still have racks that overheat because the coil nearest them is starved of flow.

> **Principle 7.3.** *Cooling capacity is a property of the network, not of a nameplate. The heat a facility can reject is limited by the weakest link along the path from rack to atmosphere, with every rating expressed on the same basis and evaluated at the present conditions.*

7.5.3 Redundancy under hot weather

Plants are built with spare units so that one failure does not reduce capacity. Take an illustrative plant that needs $N = 3$ chillers to carry a 10 MW load, each rated 3.5 MW at design conditions, and has four installed, an $N+1$ arrangement. With all four in service the plant can carry 14 MW. If one chiller fails, three remain, with $3 \times 3.5 = 10.5$ MW, still above the load: the failure costs no capacity, which is the purpose of the spare.

The rating, however, is a rating at design conditions. If the outdoor temperature rises at the same time, every chiller has a higher lift and less capacity. With an illustrative 10 % derating on a hot afternoon, each chiller carries 3.15 MW; four carry 12.6 MW, enough, but three carry only 9.45 MW, 0.55 MW short of the load. The plant is "$N+1$" by its equipment count and not by its capacity on the day that matters. Failures and hot weather are not independent either:

hot days load the equipment hardest, so they are also the days on which failures are most likely. A deterministic redundancy label describes equipment counts; it does not guarantee that the remaining equipment can carry the load under the conditions in which it will be needed. The general treatment of redundancy and reliability is the subject of a later chapter.

7.5.4 Probabilistic capacity

Every argument of the capacity relationship is uncertain. The available capacity depends on the outdoor temperature, which is forecast, not known; on the state of the equipment, which may include a unit that is about to fail; on fouling, which is inferred, not measured; on the water flows; and on model parameters θ, such as the COP curve of each chiller, that are known only approximately:

$$\dot{Q}_{\text{avail}} = f(T_{\text{amb}},\ \text{equipment state},\ \text{fouling},\ \dot{m},\ \theta). \tag{7.18}$$

With uncertain arguments, $\dot{Q}_{\text{avail}}$ is a random quantity, and the operational question is not whether the plant has enough capacity but how likely it is not to,

$$\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}}). \tag{7.19}$$

As an illustration, if the available capacity tomorrow afternoon is judged to be normally distributed with mean 11 MW and standard deviation 0.6 MW, and the required capacity is 10 MW, the probability of a shortfall is $\Phi\big((10-11)/0.6\big) = \Phi(-1.67) \approx 0.048$, about one chance in twenty, where Φ is the standard normal distribution function. A plant with a nominal margin of 1 MW is therefore not certainly safe; whether one chance in twenty is acceptable is a decision, and the probability is what makes the decision explicit.

> **Principle 7.4.** *Available cooling capacity is a random quantity that depends on the weather, the equipment state and degradation. The operational question is the probability that it falls below the required capacity, not whether a deterministic rating exceeds the load.*

7.5.5 Fouling and degradation

Most loss of cooling capacity is not a failure but a drift. Coils collect dust; air filters clog and raise the pressure the fans must overcome; heat-exchanger tubes scale; cooling-tower fill fouls with deposits and biological growth; compressors lose efficiency as valves and bearings wear; pump impellers erode. Each of these lowers a conductance or an efficiency gradually, while the equipment keeps running. As an illustrative case, a heat exchanger whose design conductance is $UA = 100$ (in any consistent unit) falls to $UA = 80$: the exchanger still runs and still transfers heat, but its maximum heat at a given driving difference falls by 20 % by equation (7.3), and by somewhat less at fixed inlet temperatures, as the coil example of Section 7.2.3 showed, 13 % in that case.

The consequence for operations is that capacity is a continuous state, not a binary one. A unit that is "available" may be available at 80 % of its rating, and the redundancy calculation of Section 7.5.3 should use the degraded value. Tracking the inferred conductance of each coil, the inferred COP of each chiller at its measured lift, and the approach of each tower over time turns degradation into a trend that can be extrapolated, so that cleaning or repair is scheduled before the capacity loss matters. This is predictive maintenance: acting on the inferred state of the equipment rather than on a fixed calendar or on a failure.

7.6 Measurement, diagnosis, control and weather

This section turns from what the equipment does to what can be known and done about it in operation: what is measured at each unit, which states remain hidden, how a symptom is diagnosed and a balance checked, how the plant is controlled, and how the weather enters as a disturbance that makes the capacity margin a forecast with a probability.

7.6.1 What is measured at each unit

A typical plant measures the following at each family of equipment.

- **CRAH**: air temperature entering and leaving the coil, $T_{\text{air,in}}$ and $T_{\text{air,out}}$; water temperature entering and leaving, $T_{w,\text{in}}$ and $T_{w,\text{out}}$; fan speed; and the position of the chilled-water valve.
- **Chiller**: chilled-water supply and return temperatures $T_{\text{CHW,s}}$ and $T_{\text{CHW,r}}$; condenser-water supply and return temperatures $T_{\text{CW,s}}$ and $T_{\text{CW,r}}$; compressor electrical power; chilled-water flow (and sometimes condenser-water flow); and refrigerant pressures in the evaporator and condenser.
- **Cooling tower**: water temperature entering and leaving, the outdoor dry-bulb temperature T_{amb} and relative humidity RH, and fan speed.

The list shows the book's three kinds of quantity at once. Temperatures, fan speeds, valve positions, compressor power and the plant flow are *measured*. The heats $\dot{Q}_{\text{air}}$, $\dot{Q}_{\text{water}}$, $\dot{Q}_{\text{evap}}$ and $\dot{Q}_{\text{cond}}$, the COP, and the wet-bulb temperature are *calculated* from measurements by known relations. Several quantities that matter are not on the list at all: the air flow through each CRAH is usually not measured but calculated from fan speed and a fan curve, an assumption that fails if a filter clogs; the water flow through each individual coil is rarely measured; and the refrigerant pressures give the evaporating and condensing temperatures, and so the lift, only through the refrigerant's saturation properties.

7.6.2 Hidden states

Some of the quantities that decide the plant's behavior are not measured and cannot be calculated directly from a single reading: the conductance UA_{coil} of each coil; the efficiency of each chiller relative to its design curve, η_{chiller}; the fouling level of coils, condensers and tower fill; the true efficiency of each pump, as distinct from its catalogue value; and the maximum available capacity $\dot{Q}^{\max}_{\text{avail}}$ under the present conditions. These are hidden states. They change slowly, they decide the capacity, and they are what maintenance acts on.

They are candidates for inference. Given measurements y, the temperatures, flows and powers of Section 7.6.1, and the relations of the earlier sections, a posterior distribution such as $p(UA \mid y)$ states what the data say about the conductance of a coil and how precisely. Propagated through the capacity relationship, it gives the quantity an operator needs,

$$\Pr(\dot{Q}_{\text{capacity}} < \dot{Q}_{\text{req}} \mid y), \tag{7.20}$$

the probability, conditioned on what has been measured, that the plant cannot carry the required load. The inference is only as good as the model and the measurements that constrain it: a coil whose water flow is not measured has a conductance that is confounded with its flow, and separating the two takes either a flow measurement or a period in which the flow changes in a known way.

7.6.3 Diagnosis of a rising return temperature

Suppose the chilled-water return temperature rises, for example from 12 to 13.5°C. The symptom is one number; the causes are several. Five hypotheses are:

- H_1: the IT load increased;
- H_2: the chilled-water flow decreased;
- H_3: a CRAH valve has a problem;
- H_4: the chiller's supply temperature increased;
- H_5: the return temperature sensor is biased.

The relations among the quantities decide between them. From equation (7.2), the return temperature is $T_{\mathrm{CHW,r}} = T_{\mathrm{CHW,s}} + \dot{Q}/(\dot{m}_w c_{p,w})$, so a rise can come from a larger heat, a smaller flow, a higher supply temperature, or a wrong reading. Each hypothesis predicts a different pattern in the other measurements. Under H_1, the loop difference grows from 5 to 6.5 K at constant flow, which means 30 % more heat, and the electrical meters on the IT load must show about 30 % more power. Under H_2, the heat is unchanged and the flow has fallen to $5/6.5 = 77\,\%$ of its value; the IT power is flat, and the pump speed, the differential pressure or the flow meter shows the change. Under H_3, a valve stuck partly closed starves one CRAH: that unit's leaving water is hotter and its supply air warmer, while the others are normal, so the signature is local. Under H_4, the supply temperature has risen by 1.5 K as well, the loop difference is unchanged at 5 K, and the cause is at the chiller, a setpoint change or a chiller at its capacity limit. Under H_5, only one sensor moved: the supply temperature, the flow, the IT power and the air-side temperatures are all unchanged, and the step may appear without the lag that a real thermal change would show.

One symptom has several causes, and the cause is identified not by the symptom but by its consistency with the rest of the measurements. The relationships among P_{IT}, $\dot{m}_w$, $T_{\mathrm{CHW,s}}$, $T_{\mathrm{CHW,r}}$, the valve positions and the equipment state are what turn a single alarm into a ranked list of explanations, each with a probability given the data, and a later chapter develops the procedure in general.

7.6.4 The three-way balance check

The agreement principle of Section 7.2.2 extends across the whole chain. As an illustrative example, the air-side calculation over all CRAHs gives $\dot{Q}_{\mathrm{air}} = 8.0\,\mathrm{MW}$, the chilled-water calculation at the plant gives $\dot{Q}_{\mathrm{CHW}} = 7.9\,\mathrm{MW}$, and the evaporator estimate built into the chiller controls gives $\dot{Q}_{\mathrm{evap}} = 6.5\,\mathrm{MW}$. In steady state, equation (7.16) says all three should be equal.

Figure 7.7 shows the three estimates as bars. The horizontal axis lists the three calculations in the order the heat passes through the chain; the vertical axis is heat in megawatts. Each bar carries an error bar of an illustrative $\pm 0.3\,\mathrm{MW}$, about the uncertainty that typical flow and temperature sensors give for a heat of this size. The dashed line is the mean of the first two estimates. The air-side and chilled-water estimates differ by 0.1 MW, about 1 %, and their error bars overlap: they agree. The evaporator estimate is 1.4 MW below the chilled-water value, about 18 %, and its error bar is far from the other two: it is inconsistent.

The inconsistency says that something is wrong, not what. The candidates are:

- a *flow-sensor bias*: the chiller's flow meter reads about 18 % low;
- a *temperature-sensor bias*: on a 5 K loop difference, an error of 0.9 K in the difference between the chiller's own supply and return sensors produces an 18 % error in the heat;

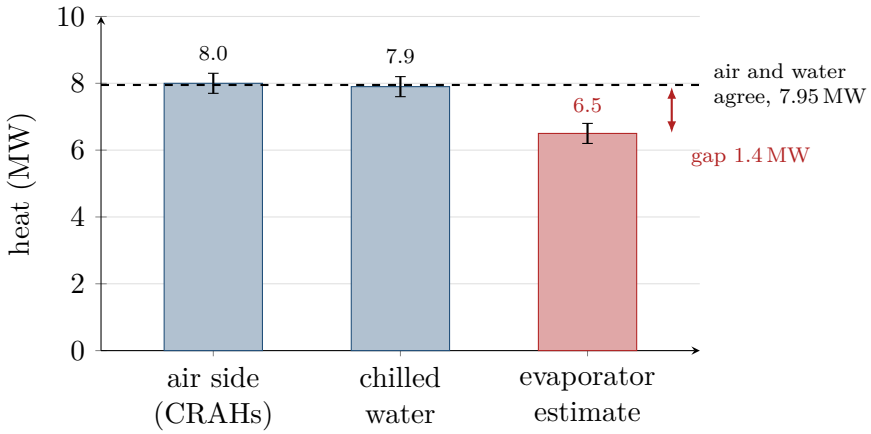


Figure 7.7: The three-way balance check, illustrative. Three independent calculations of the same heat, with error bars of ±0.3 MW. The air-side and chilled-water estimates agree within their uncertainty; the evaporator estimate falls 1.4 MW short and is inconsistent with both.

- a *missing chiller*: a second chiller is running and carrying 1.4 MW, but its evaporator heat is not in the sum;
- an *incorrect topology*: some chilled water bypasses the metered chiller through a decoupler pipe, or one CRAH is served by a different plant, so the three calculations do not cover the same flow;
- *transient storage*: the loop is warming or a thermal storage tank is charging; for an illustrative loop of 500 m^3, a 1.4 MW imbalance warms the water by $1400/(500\,000 \times 4.18) \times 3600 = 2.4$ K per hour, a change the temperature trends would show;
- a *calculation mistake*: a wrong specific heat, a unit error, or a sign convention reversed.

Each candidate makes a different prediction elsewhere, as the hypotheses of Section 7.6.3 did. The essential point is that the conservation equations detect the inconsistency automatically: nobody has to suspect the chiller's flow meter in advance, because three calculations that must agree and do not are a residual that flags itself.

7.6.5 Controls across the plant and their optimization

Each equipment family has its own actuators. Collected in one vector,

$$u = \left[u_1,\ u_2,\ u_3,\ u_4,\ u_5\right]^{\mathsf{T}}, \tag{7.21}$$

they are: u_1, the CRAH fan speed; u_2, the CRAH chilled-water valve position; u_3, the chilled-water pump speed; u_4, the chiller compressor capacity (or its supply-temperature setpoint); and u_5, the cooling-tower fan speed. The controller's task is to meet the temperature requirements of the IT equipment while minimizing energy and keeping enough capacity margin for disturbances and failures.

The controls are coupled through the chain. Raising the chilled-water supply temperature through u_4 lowers the lift and the compressor power, but the coils then need more water, u_2 and u_3, or more air, u_1, to deliver the same heat, and fan and pump power rise. Running the tower fans faster through u_5 costs fan power and returns colder condenser water, which lowers the lift and the compressor power. Each local optimum is paid for somewhere else in the plant,

so the plant is optimized as a whole. A simplified formulation is

$$\begin{aligned} \min_{u} \quad & P_{\text{fans}} + P_{\text{pumps}} + P_{\text{chillers}} + P_{\text{towers}} \\ \text{subject to} \quad & T_{\text{rack},i} \le T_{\max} \;\; \forall i, \qquad \dot{Q}_{\text{avail}} \ge \dot{Q}_{\text{req}}, \end{aligned} \tag{7.22}$$

where the objective is the total electrical power of the cooling equipment, the first constraint keeps every rack inlet below its limit, and the second keeps the chain's capacity above the requirement. When the rack temperatures are uncertain, because the loads, the airflow and the sensors are, the hard temperature constraint is replaced by a chance constraint,

$$\Pr\big(T_{\text{rack},i} > T_{\max}\big) < \epsilon \quad \forall i, \tag{7.23}$$

which allows a small probability ϵ of exceeding the limit and makes the problem one of stochastic optimization. A smaller ϵ buys safety with energy: the controller must keep a larger margin and run the plant less efficiently. Chapter 10 develops the controllers that implement such a policy.

7.6.6 Weather, the capacity margin and the risk question

The weather enters the plant directly as a disturbance: an input that acts on the system and is not chosen by the controller. Its components are the outdoor dry-bulb temperature $T_{\text{amb}}(t)$, the wet-bulb temperature $T_{\text{wb}}(t)$ and the relative humidity $\text{RH}(t)$, all functions of time. Through the tower and the lift they set the chillers' COP at each moment, $\text{COP}(t)$, the maximum cooling capacity $\dot{Q}_{\text{cool,max}}(t)$ of equation (7.15), and the cooling power $P_{\text{cool}}(t)$ needed for a given load. Because the weather can be forecast, so can these. The quantity to forecast is the thermal capacity margin at a future time $t + h$,

$$M_Q(t+h) = \dot{Q}_{\text{avail}}(t+h) - \dot{Q}_{\text{req}}(t+h), \tag{7.24}$$

the difference between the capacity the plant will have and the heat it will have to remove.

Figure 7.8 shows how the margin depends on the weather for an illustrative plant. The horizontal axis is the outdoor wet-bulb temperature; the vertical axis is cooling capacity in megawatts. The solid curve is the available capacity with all equipment in service: a constant 12 MW up to a wet-bulb temperature of 18°C, then falling by 0.25 MW per kelvin as the towers and the lift degrade. The dashed curve is the same plant with 1 MW less capacity, for example one chiller derated by fouling or out of service. The horizontal line is the required capacity, 9.5 MW. The vertical distance between the available curve and the requirement is the margin M_Q, marked by the arrow at $T_{\text{wb}} = 24$°C, where it is $10.5 - 9.5 = 1.0$ MW. The margin reaches zero at $T_{\text{wb}} = 28$°C for the full plant and at 24°C for the degraded one. The gray band is the forecast range of tomorrow afternoon's wet-bulb temperature, 26 ± 3°C, two standard deviations of an illustrative normal forecast with mean 26°C and standard deviation 1.5 K.

The risk question. The operationally useful statement is a probability,

$$\Pr\big(M_Q < 0 \mid \text{weather forecast},\ \text{current equipment state}\big), \tag{7.25}$$

the probability that cooling capacity becomes insufficient, for example tomorrow afternoon. For the full plant of Figure 7.8, the margin is negative when $T_{\text{wb}} > 28$°C, and under the forecast that probability is $1 - \Phi\big((28-26)/1.5\big) = 1 - \Phi(1.33) \approx 0.09$. For the degraded plant the threshold is 24°C and the probability is $1 - \Phi(-1.33) \approx 0.91$: the same weather that is a modest risk for the full plant is a near certainty of shortfall for the degraded one. Two further questions follow from the same calculation. The first is which uncertain component contributes most to the risk: here the equipment state matters more than the weather spread, since a 1 MW change in capacity moves the threshold by 4 K, more than two forecast standard deviations. The second

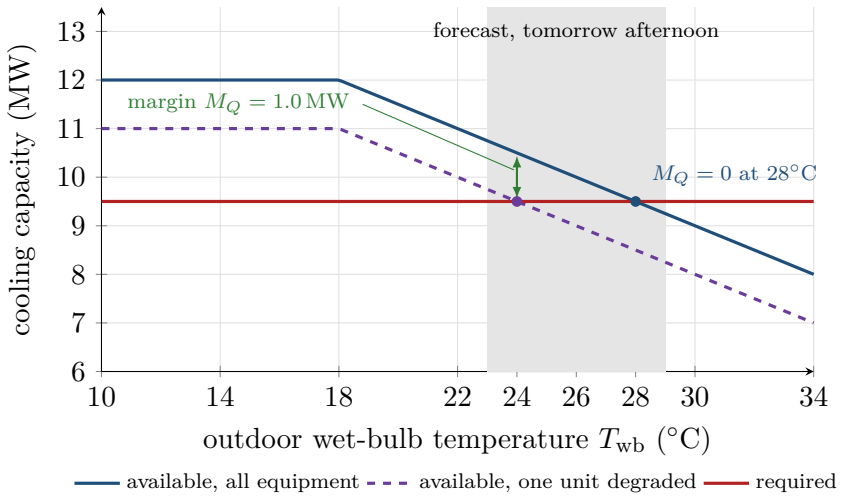


Figure 7.8: Available cooling capacity against outdoor wet-bulb temperature, from the illustrative closed form $\dot{Q}_{\text{avail}} = 12 - 0.25\max(0, T_{\text{wb}} - 18)$ MW (solid) and the same minus 1 MW for one degraded unit (dashed), against a required capacity of 9.5 MW. The margin M_Q is the vertical gap; the gray band is a forecast of 26 ± 3°C (± 2 standard deviations).

is which action reduces the risk most: returning the degraded unit to service, cleaning a fouled coil, starting the spare chiller in advance, postponing planned maintenance, precooling the building and the water loops before the afternoon, or moving IT load elsewhere. Each action shifts a curve of Figure 7.8 or the requirement, and its value is the reduction in probability it buys.

7.6.7 The equations of the chain

Five relations carry most of the chapter:

$\dot{Q} = \dot{m}\, c_p\, \Delta T$	sensible heat carried by a stream, eqs. (7.1), (7.2);
$\dot{Q} \approx UA\, \Delta T$	heat a heat exchanger can pass, eq. (7.3);
$P_{\text{pump}} = \Delta p\, \dot{V}/\eta$	electrical power to move water, eq. (7.7);
$\dot{Q}_{\text{cond}} = \dot{Q}_{\text{evap}} + W_{\text{comp}}$	energy balance of a chiller, eq. (7.9);
$\text{COP} = \dot{Q}_{\text{evap}}/W_{\text{comp}}$	chiller performance, eq. (7.10).

The first states what a stream carries, the second what an exchanger allows, the third what moving the stream costs, the fourth that the chiller conserves energy, and the fifth what the chiller costs per unit of heat. Combined along the chain of Figure 7.1, with the capacity of each link and the weather as an input, they give the heat the facility can reject.

The point of the chapter is the combination. A cooling system is a chain of coupled heat-transport components, air, coil, water, chiller and heat rejection, each with a capacity, a resistance, an efficiency, controls, a possible degradation and an uncertainty in its measurements. The relevant question is therefore not whether the chiller is big enough. It is whether the complete heat-removal path can carry the required load under the current conditions and under the conditions forecast for the hours ahead.

7.7 What to retain

Definition 7.1 (Heat-rejection chain)**.** A cooling system is a chain of coupled heat-transport components, from air through coils and water to chillers and heat rejection, each with a capacity,

a resistance, an efficiency, controls, possible degradation and measurement uncertainty. Its capacity is that of the complete path under the present and forecast weather, and the operational question is the probability that this path cannot carry the required load.

Table 7.2 lists the chapter's objects and the subsection that defines each.

Table 7.2: The objects of the cooling chain and where each is defined.

Object	Expression	Role	Defined in
Reference chain	Figure 7.1	rack to atmosphere through two water loops	7.1.1
Air-side, water-side heat	equations (7.1), (7.2)	two independent calculations of the coil heat	7.2.2
Overall conductance	$\dot{Q} = UA\,\Delta T_{\mathrm{lm}}$	what a coil can pass; lowered by fouling	7.2.3
System curve	$\Delta p \approx K\dot{V}^2$	resistance of the water circuit	7.3.1
Pump power	$P_{\mathrm{pump}} = \Delta p\dot{V}/\eta$	cost of moving the carrier	7.3.2
Affinity laws	$\dot{V} \propto N_{\mathrm{pump}}$, $\Delta p \propto N_{\mathrm{pump}}^2$, $P \propto N_{\mathrm{pump}}^3$	savings from variable speed	7.3.2
Chiller balance	$\dot{Q}_{\mathrm{cond}} = \dot{Q}_{\mathrm{evap}} + W_{\mathrm{comp}}$	heat rejected exceeds heat removed	7.3.4
Coefficient of performance	$\mathrm{COP} = \dot{Q}_{\mathrm{evap}}/W_{\mathrm{comp}}$	transport ratio, not an efficiency	7.3.5
Temperature lift	$T_{\mathrm{cond}} - T_{\mathrm{evap}}$	sets the COP	7.3.6
Latent heat	$\dot{Q} = \dot{m}_{\mathrm{evap}} L_v$	how a tower rejects heat	7.4.2
Wet-bulb temperature	T_{wb}; approach	limit of evaporative cooling	7.4.3
Capacity relationship	equation (7.15)	capacity as a function of weather and state	7.4.4
Chain capacity	min of the link capacities	the bottleneck	7.5.2
Probabilistic capacity	$\Pr(\dot{Q}_{\mathrm{avail}} < \dot{Q}_{\mathrm{req}})$	the operational question	7.5.4
Hidden states	UA, η_{chiller}, fouling, $\dot{Q}_{\mathrm{avail}}^{\max}$	inferred, not measured	7.6.2
Three-way balance	air, chilled water, evaporator	residual that flags inconsistency	7.6.4
Plant optimization	equations (7.22), (7.23)	minimum energy under a chance constraint	7.6.5
Capacity margin	$M_Q = \dot{Q}_{\mathrm{avail}} - \dot{Q}_{\mathrm{req}}$	forecast with a probability of shortfall	7.6.6

7.8 Exercises

Exercise 7.1. A CRAH coil cools 40 kg/s of air from 36 to 22°C, with $c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ/(kg\,K)}$, using chilled water that enters at 8°C and leaves at 14°C. (a) Compute the air-side heat and the water flow needed to carry it. (b) The coil's flow meter reads 20 kg/s. Compute the water-side heat it implies and the percentage disagreement with the air side, and list three explanations that would account for it. (c) Treating the coil as a counterflow exchanger, compute ΔT_{lm} and UA from the air-side heat.

Exercise 7.2. A chilled-water pump delivers $0.30\,\mathrm{m^3/s}$ against 200 kPa with an overall efficiency of 0.70, on a loop whose system curve passes through the origin. (a) Compute the hydraulic and electrical power. (b) Using the affinity laws, compute flow, pressure and electrical power at 70 % speed. (c) The loop carries 5 MW. Compute the loop temperature difference at full and at 70 % speed, taking $1000\,\mathrm{kg/m^3}$ and $c_{p,w} = 4.18\,\mathrm{kJ/(kg\,K)}$, and state what the coils must achieve for the reduced flow to be acceptable.

Exercise 7.3. A chiller removes $\dot{Q}_{\text{evap}} = 4\,\text{MW}$ at COP $= 5$. (a) Compute W_{comp} and $\dot{Q}_{\text{cond}}$. (b) On a hot afternoon the COP falls to 4 at the same $\dot{Q}_{\text{evap}}$; compute the new W_{comp} and $\dot{Q}_{\text{cond}}$, and the water that a tower rejecting all of $\dot{Q}_{\text{cond}}$ as latent heat with $L_v = 2400\,\text{kJ/kg}$ would evaporate per second. (c) The tower can reject at most 4.9 MW. What is the largest evaporator heat the plant can carry at COP $= 4$, and is the 4 MW load still covered?

Exercise 7.4. A plant's available capacity is $\dot{Q}_{\text{avail}} = 12 - 0.25\max(0, T_{\text{wb}} - 18)\,\text{MW}$, as in Figure 7.8, and the required capacity is 9.5 MW. (a) Compute the margin M_Q at $T_{\text{wb}} = 24$ and 30°C, and the wet-bulb temperature at which it reaches zero. (b) Tomorrow afternoon's wet-bulb forecast is normal with mean 27°C and standard deviation 1 K. Compute $\Pr(M_Q < 0)$. (c) Starting a spare chiller adds 1 MW to the available capacity at every temperature; compute the new threshold and probability. (d) Instead, fouled coils remove 0.5 MW; compute the threshold and probability, and state which of the two changes matters more and why.

7.9 Connection to the TERRA framework

The heat-rejection chain is a network of conservation laws coupled to the electrical network through the compressors, pumps and fans, and it maps onto a TERRA graph directly.

- **Carriers are nodes, heat flows are edges.** The room air, the chilled-water loop, the refrigerant and the condenser-water loop are thermal-energy nodes, and $\dot{Q}_1$ to $\dot{Q}_5$ of Figure 7.1 are the edges between them, with equation (7.16) as the steady balance at each node and stored heat as its state in transients. Water and air mass flows are edges of a separate mass network whose pressure balance, equations (7.5) and (7.6), sets the flow that the thermal edges carry.

- **Equipment are components with laws, bounds and states.** The coil is a component with conductance UA (equation (7.3)); the pump has a curve and an efficiency; the chiller has a COP that depends on the lift and joins the thermal network to the electrical one through $\dot{Q}_{\text{cond}} = \dot{Q}_{\text{evap}} + W_{\text{comp}}$; the tower has a capacity that depends on T_{wb}. Each has a capacity bound, and the bottleneck of Figure 7.6 is the binding bound along the path.

- **Balances that should close are residuals.** The air-side against water-side agreement and the three-way check of Figure 7.7 are redundancy relations: two or three calculations of the same edge flow from different measurements. A residual beyond its uncertainty flags a sensor bias, a missing chiller or a wrong topology without anyone having to suspect it.

- **Hidden states are inferred as posteriors.** Coil conductance, chiller efficiency, fouling and pump efficiency are parameters inferred from measured temperatures, flows and powers, as $p(UA \mid y)$. The hypotheses H_1 to H_5 for a rising return temperature become a ranked posterior, and the next measurement, such as a clamp-on flow meter on one coil, is chosen for how much it separates them.

- **Risk is a probability from forward simulation.** Weather scenarios drawn from a forecast and equipment states drawn from the posterior are pushed through the graph to give the distribution of $\dot{Q}_{\text{avail}}$, and the margin risk of equation (7.25) is the fraction of scenarios with $M_Q < 0$, together with the component whose uncertainty contributes most to it.

In summary, cooling equipment is a series of heat-transport components whose balances, capacities and efficiencies form a network; its deliverable capacity is a property of the whole path under the present weather and equipment state, and that capacity, its degradation and the

faults hidden in its measurements are inferred rather than read. The next chapter adds time: the thermal mass of rooms, equipment and water loops, and how the chain responds during transients such as a chiller trip or a sudden load change.

Chapter 8

Thermal dynamics: capacitance, time constants, delays and time to limit

8.1 The dynamic heat balance

Chapter 5 wrote the heat balance of a data center in steady state: the heat removed by the cooling system equals the heat generated by the IT equipment, and the temperatures are whatever values make the two equal. That balance answers where the temperatures settle. It does not say how they get there, and operation is mostly about the getting there: a load that changes, a fan that slows, a chiller that trips. The question of this chapter is

when the load changes or the cooling degrades, how fast do the temperatures of a data center respond, and what does that response reveal about the facility?

The answer rests on one modification of the steady balance: a term for the heat that is stored. With it, temperature becomes a state that carries memory, and the chapter follows the consequences: time constants, delays, the propagation of a disturbance through the facility, the interaction of control loops, the time that remains before a limit is reached, and what a temperature record can determine about the thermal mass and resistances that produce it.

Jargon: thermal capacitance, thermal inertia, time constant, step response, transport delay, ride-through, time to limit, system identification

Thermal capacitance C_{th}, in joules per kelvin, is the heat needed to raise the temperature of a body by one kelvin; for a body of mass m and specific heat c_p it is $m\,c_p$. *Thermal inertia* is the resulting reluctance of a body with large thermal capacitance to change temperature quickly. The *time constant* τ of a first-order system is the time its response takes to cover about 63 % of a step toward its new equilibrium; for a thermal node it is the product of a thermal resistance and a thermal capacitance. A *step response* is the trajectory of an output after an input changes abruptly from one constant value to another and then stays there. A *transport delay* (also dead time) is the time a fluid takes to carry a change from one place to another; during it the downstream point shows nothing at all. *Ride-through* is the ability of a system to keep operating within limits for a period after a supply fails, here the cooling supply, by drawing on stored capacity. The *time to limit* is the time until a temperature, on its present trajectory, reaches its allowed maximum. *System identification* is the estimation of a dynamic model's parameters from measured inputs and outputs, usually from a deliberately or naturally excited transient.

Key idea

A data center's temperatures are states of a heat balance with storage, $C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}}$: an imbalance changes the rate of temperature, not the temperature itself. The operational question therefore moves from whether the facility is within limits now to where, given its state and its dynamics, it is heading, and how long it takes to get there.

8.1.1 The balance with storage

Consider a control volume, a room, an aisle, a rack, or a loop of chilled water, with a single representative temperature T. Heat enters it at a rate $\dot{Q}_{\text{in}}$, the heat of the IT equipment or of the air arriving from upstream, and leaves at a rate $\dot{Q}_{\text{out}}$, the heat carried off by the cooling system. Energy conservation for the volume reads

$$C_{\text{th}}\,\frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}}, \tag{8.1}$$

where the left-hand side is the rate at which heat is stored in the volume's thermal mass. Term by term,

$$\text{rate of heat stored} = \text{heat in} - \text{heat out}.$$

The sign of the right-hand side sets the direction of the temperature. If more heat enters than leaves, $\mathrm{d}T/\mathrm{d}t > 0$ and the temperature rises; if less, it falls; if the two are equal, the temperature holds. The steady balance of Chapter 5 is the third case, the special case $\mathrm{d}T/\mathrm{d}t = 0$ of equation (8.1).

The equation fixes the *rate of change* of temperature, not its value. A heat imbalance does not set a new temperature at once; it sets a slope, and the temperature reached after some time is the integral of that slope. This is the whole content of thermal dynamics, and every section of the chapter is a reading of it: what makes up $\dot{Q}_{\text{out}}$, how large C_{th} is, how many such volumes are connected, and how long the heat takes to travel between them.

8.1.2 Thermal capacitance

The thermal capacitance of a body is the energy needed to change its temperature by one kelvin,

$$C_{\text{th}} = m\,c_p, \tag{8.2}$$

in joules per kelvin, with m the mass in kilograms and c_p the specific heat in joules per kilogram per kelvin. A composite volume, such as a room with its air, its racks and its floor, has the sum of the capacitances of its parts, $C_{\text{th}} = \sum_i m_i c_{p,i}$, provided the parts are close enough in temperature to be treated as one.

An illustrative capacitance of $C_{\text{th}} = 10\,\text{MJ/K}$ means that 10 MJ of stored heat raises the temperature by about 1°C, and 20 MJ by about 2°C. For scale, 10 MJ is $10/3.6 \approx 2.8$ kWh, which a 1 MW IT load deposits in ten seconds. Capacitance is the thermal analog of the size of a tank: the same inflow raises the level of a small tank quickly and of a large tank slowly. Increments of temperature are the same in kelvin and in degrees Celsius, so a capacitance in J/K is equally a capacitance in joules per degree Celsius.

8.1.3 Thermal inertia

If the capacitance were zero, equation (8.1) would force $\dot{Q}_{\text{in}} = \dot{Q}_{\text{out}}$ at every instant, and the temperature would jump to whatever value made the two equal the moment either changed. No physical volume behaves that way, because every part of a data center stores thermal energy: the processors and heat sinks of the servers, the metal of the chassis and racks, the room air,

the concrete of the floor and walls, the water in the chilled-water loop, the piping, and the heat exchangers of the cooling units. Together they give the facility thermal inertia: its temperatures lag the imbalances that drive them.

The parts differ widely in how much they store. The air of an illustrative hall of 3000 m^2 with a 4 m ceiling occupies 12 000 m^3, about 14 400 kg at 1.2 $\mathrm{kg/m}^3$, and with $c_{p,\mathrm{air}} \approx 1.005\,\mathrm{kJ/(kg\,K)}$ its capacitance is only about 14.5 MJ/K. The metal of several hundred racks of servers, the floor slab and the water in the cooling loops each hold several times more. Air is the medium that moves the heat, but it is not where most of the heat is stored; the effective capacitance that governs a room's response is dominated by solids and water, coupled to the air through thermal resistances, which is why Section 8.3 replaces the single volume by several.

8.1.4 A cooling shortfall

Take an illustrative hall whose IT equipment dissipates $\dot{Q}_{\mathrm{IT}} = 1\,\mathrm{MW}$ and whose cooling, after a partial failure, can remove only 0.8 MW, its maximum with the remaining units. The imbalance is $0.2\,\mathrm{MW} = 2 \times 10^5\,\mathrm{W}$. With an effective thermal capacitance of $C_{\mathrm{th}} = 100\,\mathrm{MJ/K} = 10^8\,\mathrm{J/K}$, equation (8.1) gives

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{2 \times 10^5\,\mathrm{W}}{10^8\,\mathrm{J/K}} = 0.002\,\mathrm{K/s} = 0.002 \times 60 = 0.12^{\circ}\mathrm{C/min},$$

so after ten minutes the temperature has risen by about $0.12 \times 10 = 1.2^{\circ}\mathrm{C}$. The rise is linear because the cooling is at its maximum: its heat removal no longer grows when the room warms, so the imbalance, and with it the slope, stays constant. Operators have time to react: a little over a degree in ten minutes leaves room to start a standby unit, shed load or open a bypass before any limit is reached. Had the capacitance been only that of the room air, 14.5 MJ/K, the same shortfall would have raised the temperature at $2 \times 10^5 / 1.45 \times 10^7 = 0.0138\,\mathrm{K/s}$, about $0.83^{\circ}\mathrm{C/min}$, and the margin for action would have been seven times shorter.

Figure 8.1 draws the example. The horizontal axis is minutes since the shortfall began; the vertical axis is the hall temperature, starting from an illustrative 24°C. The solid line is the case $C_{\mathrm{th}} = 100\,\mathrm{MJ/K}$, with its value after ten minutes marked; the dashed and dotted lines are the same 0.2 MW shortfall with half and twice that capacitance. All three are straight lines, and their slopes are in inverse proportion to the capacitance: thermal mass does not prevent the rise, it buys time.

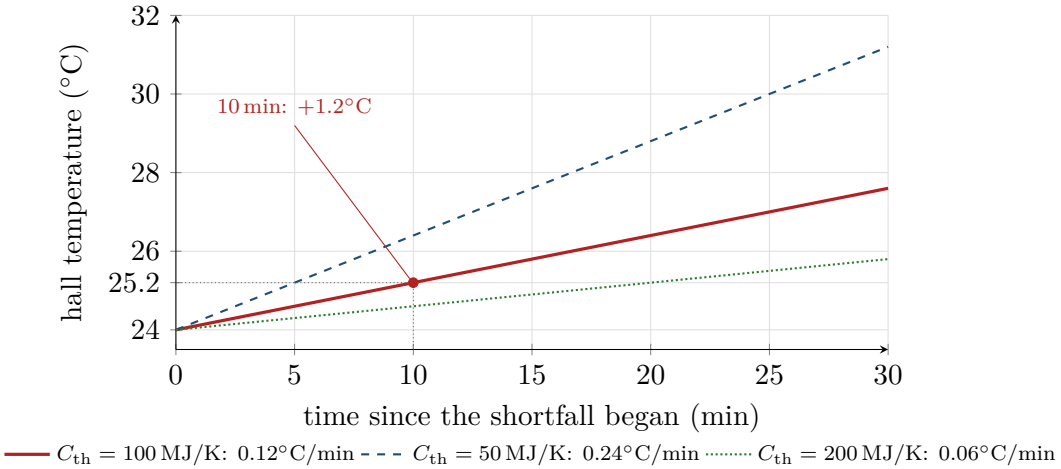


Figure 8.1: Hall temperature under a constant cooling shortfall of 0.2 MW ($\dot{Q}_{\mathrm{IT}} = 1\,\mathrm{MW}$, cooling at its maximum of 0.8 MW), $T(t) = 24^{\circ}\mathrm{C} + (0.2\,\mathrm{MW}/C_{\mathrm{th}})\,t$, for three illustrative effective capacitances. The slope is the imbalance divided by the capacitance.

8.2 The first-order thermal model

The linear rise of Section 8.1.4 holds only while the heat removal is fixed. In most situations removal grows as the volume warms, and the temperature then approaches an equilibrium instead of rising without end. This section writes that model, its equilibrium, its time constant and its step response.

8.2.1 Heat loss that grows with temperature

Temperature does not rise forever, because the heat that leaves a warm volume depends on how warm it is. Heat flows from hot to cold, and across a given path its rate is, to a first approximation, proportional to the temperature difference that drives it. For a volume at temperature T cooled by a medium at temperature T_{cool}, the supply air of a cooling unit or the water in a coil,

$$\dot{Q}_{\text{out}} = \frac{T - T_{\text{cool}}}{R_{\text{th}}}, \tag{8.3}$$

where R_{th} is the thermal resistance of the path. Substituting into equation (8.1) with the IT heat as the input gives

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{IT}} - \frac{T - T_{\text{cool}}}{R_{\text{th}}}, \tag{8.4}$$

the *first-order thermal model*. It is called first order because it contains a single first derivative of a single state. Its behavior is self-correcting: as T rises, the loss term grows, the imbalance shrinks, and the rise slows. The constant-slope case of Section 8.1.4 is the limit in which the cooling is saturated and the loss no longer depends on T.

Thermal resistance. The thermal resistance R_{th}, in kelvin per watt, is the temperature difference needed to drive one watt of heat across a path:

$$\dot{Q} = \frac{\Delta T}{R_{\text{th}}}. \tag{8.5}$$

It is the thermal analog of Ohm's law, with temperature difference in place of voltage and heat flow in place of current. A high resistance means poor heat transfer: a large temperature difference is needed to move a given heat. Low airflow, a fouled coil, a blocked filter or a heat sink with a poor contact all raise the resistance of the path they sit on; more airflow lowers it. Its inverse $1/R_{\text{th}}$, in watts per kelvin, is a conductance, the quantity written UA for a heat exchanger in Chapter 7.

8.2.2 Equilibrium temperature

Setting $\mathrm{d}T/\mathrm{d}t = 0$ in equation (8.4) gives the temperature at which the loss exactly matches the input,

$$T_{\text{eq}} = T_{\text{cool}} + R_{\text{th}}\,\dot{Q}_{\text{IT}}. \tag{8.6}$$

The volume settles above the cooling medium by the heat it must reject times the resistance of the path. For an illustrative rack zone with $T_{\text{cool}} = 20\,^{\circ}\text{C}$, $R_{\text{th}} = 0.2\,\text{K/kW}$ and $\dot{Q}_{\text{IT}} = 20\,\text{kW}$,

$$T_{\text{eq}} = 20\,^{\circ}\text{C} + 0.2\,\text{K/kW} \times 20\,\text{kW} = 20 + 4 = 24\,^{\circ}\text{C}.$$

Doubling the load to 40 kW doubles the rise to 8 K, and the zone settles at 28°C; doubling the resistance, for example by halving the effective airflow, has the same effect. The capacitance does not appear in equation (8.6): thermal mass decides how long the volume takes to reach equilibrium, not where the equilibrium is.

8.2.3 The time constant

Write the deviation from equilibrium as $\theta = T - T_{\text{eq}}$ for a constant input. Since T_{cool} and $\dot{Q}_{\text{IT}}$ are constant, $\mathrm{d}\theta/\mathrm{d}t = \mathrm{d}T/\mathrm{d}t$, and equation (8.4) becomes

$$R_{\text{th}}C_{\text{th}}\,\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\theta, \qquad \tau = R_{\text{th}}C_{\text{th}}. \tag{8.7}$$

The product τ has units of (K/W)(J/K) = J/W = s; it is the time constant of the volume. The deviation decays as $\theta(t) = \theta(0)\,e^{-t/\tau}$, so after a step the fraction of the final change already completed is $1 - e^{-t/\tau}$:

elapsed time	1τ	2τ	3τ	5τ
fraction of the final change	63.2 %	86.5 %	95.0 %	99.3 %

After five time constants the response is effectively settled. A large capacitance gives a slow response, and so does a large resistance, because a poorly cooled volume both settles higher and gets there more slowly.

A ten-minute time constant. Take an illustrative rack and its immediate air, with $R_{\text{th}} = 0.002\,\text{K/W}$ and $C_{\text{th}} = 300\,000\,\text{J/K}$. Then

$$\tau = 0.002\,\text{K/W} \times 300\,000\,\text{J/K} = 600\,\text{s} = 10\,\text{min}.$$

A load increase of 2 kW on this rack raises its equilibrium by $0.002 \times 2000 = 4\,\text{K}$, but after ten minutes only 63 % of that, 2.5 K, has appeared; 95 % appears after half an hour and the response is settled after about fifty minutes. A load change takes tens of minutes to express itself fully in temperature, and a reading taken a few minutes after the change shows only a fraction of its eventual effect.

8.2.4 The step response

For a constant input switched on at $t = 0$, starting from a temperature T_0, the solution of equation (8.4) is

$$T(t) = T_\infty + (T_0 - T_\infty)\,e^{-t/\tau}, \qquad T_\infty = T_{\text{cool}} + R_{\text{th}}\,\dot{Q}_{\text{IT}}, \tag{8.8}$$

which can be checked by differentiation: $\mathrm{d}T/\mathrm{d}t = -(T_0 - T_\infty)e^{-t/\tau}/\tau = (T_\infty - T)/\tau$, which is equation (8.4) divided by C_{th}. Two features of the solution matter in practice. The temperature does not jump: it is continuous at $t = 0$, whatever the input does. And it approaches its final value asymptotically, quickly at first and ever more slowly, never quite arriving. Its initial slope is $(T_\infty - T_0)/\tau$; a straight line with that slope reaches the final value exactly at $t = \tau$, which gives a second way to read the time constant off a measured curve.

Figure 8.2 draws the response of the ten-minute rack of Section 8.2.3 to the 2 kW step. The horizontal axis is minutes since the step; the vertical axis is the rack temperature, starting from an illustrative 24°C. The dashed horizontal line is the final value $T_\infty = 28$°C. The dots mark the temperature at one, two, three and five time constants, labeled with the fraction of the final change completed; the dotted line from the origin, labeled tangent, is the initial slope. Most of the change happens in the first two time constants; the last few percent take as long again.

> **Principle 8.1.** *A thermal volume responds to a change of input with a time constant $\tau = R_{\text{th}}C_{\text{th}}$: resistance and input set where the temperature goes, $T_\infty = T_{\text{cool}} + R_{\text{th}}\dot{Q}_{\text{IT}}$, and capacitance together with resistance sets how fast it gets there.*

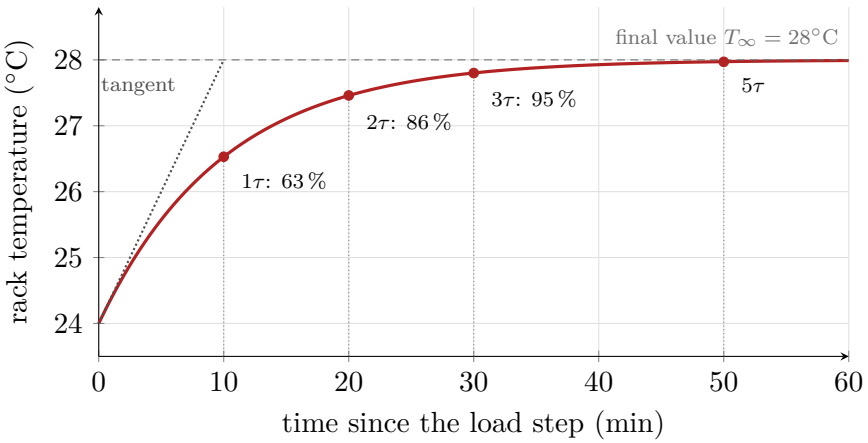


Figure 8.2: First-order step response, $T(t) = T_\infty + (T_0 - T_\infty)e^{-t/\tau}$ (equation (8.8)), with illustrative $T_0 = 24°C$, $T_\infty = 28°C$ (a 2 kW step through $R_{th} = 0.002\,K/W$) and $\tau = R_{th}C_{th} = 10\,min$ ($C_{th} = 300\,kJ/K$). The response is continuous, completes 63, 86 and 95 % of the change at one, two and three time constants, and is settled after about five.

8.3 Memory, multiple masses and storage

A single first-order volume already gives temperature a memory of past inputs. This section draws the consequences for interpreting and modeling temperature records, replaces the single volume by several coupled masses with different speeds, and treats the largest store of all, the chilled water, as a reserve for ride-through.

8.3.1 The lag between power and temperature

Electrical power and temperature respond on different time scales. When a batch of jobs starts, the IT power P_{IT} rises within milliseconds to seconds, and since the electrical power of the IT equipment is converted to heat as it is consumed, $\dot{Q}_{IT} = P_{IT}$ rises with it (Chapter 5). The temperatures do not jump; they follow gradually, with the time constants of Section 8.2.3. In words: a change in P_{IT} at time t shows in the temperature T at a later time $t + \Delta t$, and the lag Δt is not a single number but a spread, because the response is an exponential approach rather than a shifted copy of the input.

The lag matters wherever a temperature is interpreted. A temperature spike observed now may have been caused by a load change several minutes earlier, at a moment when the temperature showed nothing. For diagnosis this means that the cause of a temperature excursion is looked for in the inputs over the preceding few time constants, not in the inputs of the same instant. For forecasting it means that the present temperature does not summarize the future: a rack whose load rose two minutes ago has most of its temperature rise still to come.

8.3.2 Lagged inputs in data-driven models

A data-driven model that relates temperature only to the power of the same time step,

$$T_t = f(P_t),$$

performs poorly, because the physical system has memory and the present temperature depends on past power. A model with lagged inputs,

$$T_t = f(P_t, P_{t-1}, P_{t-2}, \dots),$$

performs better, and the first-order model says exactly how. If the heat input is constant over each sampling interval Δt, the solution of equation (8.4) over one interval is equation (8.8) with $t = \Delta t$, which gives the exact recursion

$$T_{k+1} = a\,T_k + (1-a)(T_{\text{cool}} + R_{\text{th}}\,\dot{Q}_k), \qquad a = e^{-\Delta t/\tau}. \tag{8.9}$$

Unrolling it n steps back,

$$T_k = a^n\,T_{k-n} + (1-a)\sum_{j=1}^{n} a^{j-1}(T_{\text{cool}} + R_{\text{th}}\,\dot{Q}_{k-j}), \tag{8.10}$$

shows that the present temperature is a weighted sum of past heat inputs with weights that decay geometrically with age. With one-minute samples and the ten-minute time constant of Section 8.2.3, $a = e^{-0.1} = 0.905$: the most recent minute carries a weight of $1 - a = 0.095$, the minute ten steps back a weight of $0.095 \times 0.905^9 \approx 0.039$, the last ten minutes together $1 - a^{10} = 0.63$ of the total and the last thirty minutes 0.95. A model that uses only the present power uses less than a tenth of the information that determines the present temperature.

Physical memory. In a recurrent neural network, memory is computational: a hidden vector carried from step to step, whose meaning the training must discover. In a thermal system the memory exists physically, as heat stored in a thermal capacitance C_{th}, and equation (8.10) gives its form: one decay factor per stored mass, set by R_{th} and C_{th}. Writing $C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t$ into the model explicitly, instead of leaving a black-box model to discover the lag behavior from data, replaces a free weight for every lag by two physical parameters, constrains the weights to the shape physics allows, and makes the fitted model interpretable: its parameters are a resistance and a capacitance that can be compared with the equipment.

8.3.3 Multiple thermal masses

A data center does not have one temperature. Along the path of the heat there are several masses, each with its own temperature and capacitance: the processor die and heat sink at T_{chip}, the air and metal inside the server at T_{air}, the rack, and the room. Each mass exchanges heat with its neighbors through a resistance. For the first two, with the chip dissipating $\dot{Q}_{\text{chip}}$ and the server air rejecting heat to the room at T_{room},

$$C_{\text{chip}}\,\frac{\mathrm{d}T_{\text{chip}}}{\mathrm{d}t} = \dot{Q}_{\text{chip}} - \frac{T_{\text{chip}} - T_{\text{air}}}{R_1}, \tag{8.11}$$

$$C_{\text{air}}\,\frac{\mathrm{d}T_{\text{air}}}{\mathrm{d}t} = \frac{T_{\text{chip}} - T_{\text{air}}}{R_1} - \frac{T_{\text{air}} - T_{\text{room}}}{R_2}, \tag{8.12}$$

where R_1 is the resistance from the chip to the server air and R_2 from the server air to the room. The heat leaving the chip in equation (8.11) is exactly the heat entering the air in equation (8.12): the two balances share an edge. Together they form a second-order system: two states, two time constants, and a response that is a sum of two exponentials.

For an illustrative server with $\dot{Q}_{\text{chip}} = 500\,\text{W}$, $C_{\text{chip}} = 2\,\text{kJ/K}$, $R_1 = 0.02\,\text{K/W}$, $C_{\text{air}} = 30\,\text{kJ/K}$ and $R_2 = 0.01\,\text{K/W}$, the equilibrium has the air $500 \times 0.01 = 5\,\text{K}$ above the room and the chip a further $500 \times 0.02 = 10\,\text{K}$ above the air, 15 K in all. The two uncoupled products are $R_1 C_{\text{chip}} = 40\,\text{s}$ and $R_2 C_{\text{air}} = 300\,\text{s}$; the coupled system, whose time constants are the inverse eigenvalues of the 2×2 matrix of equations (8.11)–(8.12), has time constants of 37 s and 5.4 min. After a step of the chip power from zero, with the room held constant, the two rises above the room are, with t in minutes,

$$T_{\text{chip}} - T_{\text{room}} = 15 - 8.55\,e^{-1.614\,t} - 6.45\,e^{-0.186\,t},$$
$$T_{\text{air}} - T_{\text{room}} = 5 + 0.65\,e^{-1.614\,t} - 5.65\,e^{-0.186\,t}.$$

Figure 8.3 plots these two curves. The horizontal axis is minutes since the chip power step; the vertical axis is the rise above the room temperature. The red curve is the chip, the blue curve the server air, and the dashed lines their equilibria. The chip shows both components clearly: within the first two minutes it climbs by about 10 K, mostly the fast component, as its small mass charges against the resistance R_1; it then creeps up the remaining 5 K over the next fifteen minutes, the slow component, because the air it rejects heat to is itself warming. The air shows almost only the slow component. A single-time-constant model fitted to the chip curve would fit neither part well.

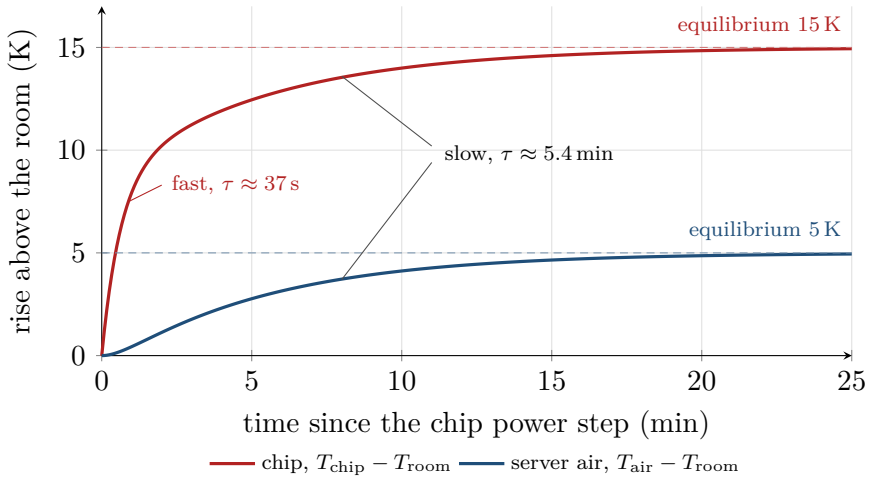


Figure 8.3: Two-mass response to a 500 W chip power step, equations (8.11)–(8.12), with illustrative $C_{\text{chip}} = 2\,\text{kJ/K}$, $R_1 = 0.02\,\text{K/W}$, $C_{\text{air}} = 30\,\text{kJ/K}$, $R_2 = 0.01\,\text{K/W}$ and the room held constant. Closed forms (t in min): chip $15 - 8.55e^{-1.614t} - 6.45e^{-0.186t}$, air $5 + 0.65e^{-1.614t} - 5.65e^{-0.186t}$. The chip shows a fast and a slow component; the air almost only the slow one.

8.3.4 Different speeds

Each mass in the chain responds at its own speed, and the speeds span several orders of magnitude. Table 8.1 lists the layers of a typical facility with the medium that stores the heat, an illustrative order of magnitude of the time constant, and the control that acts at that scale. Chips respond in seconds; the air inside servers and racks in seconds to minutes; room air coupled to equipment in minutes; the building structure and the water systems over minutes to hours.

The spread matters for control. A controller tuned for one time scale may be unsuited to another: a loop that acts every few seconds on a water temperature that takes half an hour to respond will act many times on stale information and drive itself into oscillation (Section 8.5.1); a loop that acts every fifteen minutes on a chip temperature will miss every excursion. It matters for measurement as well: a sampling interval adequate for the room is far too coarse for the chip, and a model that lumps fast and slow layers into one capacitance fits neither.

8.3.5 Chilled water as thermal storage

The water in a chilled-water system is one of the largest thermal masses in a facility. Its stored thermal energy, relative to a reference temperature, is

$$E_w = m_w\, c_{p,w}\, T_w, \qquad C_w = m_w\, c_{p,w}, \tag{8.13}$$

where m_w is the mass of water in the loop and T_w its temperature. For an illustrative loop holding $m_w = 100\,000\,\text{kg}$ of water, a hundred cubic meters in pipes, coils and buffer vessels,

Table 8.1: Thermal layers of a data center and their time scales. Orders of magnitude are illustrative; actual values depend on mass, airflow and design.

Layer	Heat stored in	Time constant	Control acting at this scale
Chip	die, heat spreader, heat sink	seconds	processor frequency throttling, server fan firmware
Server and rack air	air, chassis and rack metal	seconds to minutes	server fans; containment doors
Room or aisle	room air coupled to equipment	minutes	cooling-unit fan speed, supply-air setpoint
Building structure	floor slab, walls	hours	none directly; it buffers the room
Chilled-water loop	water in pipes and coils	minutes to hours	valve position, pump pressure, chiller supply setpoint
Thermal storage tank	stored chilled water	hours	charge and discharge scheduling

with $c_{p,w} = 4.18\,\mathrm{kJ/(kg\,K)}$,

$$C_w = 100\,000\,\mathrm{kg} \times 4.18\,\mathrm{kJ/(kg\,K)} = 418\,000\,\mathrm{kJ/K} = 418\,\mathrm{MJ/K},$$

more than four times the effective capacitance of the hall of Section 8.1.4 and nearly thirty times that of its air. Water's specific heat is about four times that of air per kilogram and its density is about eight hundred times larger, so a modest volume of water is a substantial thermal buffer.

Ride-through during a chiller outage. If a chiller stops while the pumps and the cooling-unit fans keep running, typically on uninterruptible power, the cold water already in the loop keeps absorbing heat from the coils while its temperature rises. With an illustrative 1 MW of heat entering the loop and the 418 MJ/K of the example, the loop warms at $10^6/4.18 \times 10^8 = 0.0024\,\mathrm{K/s}$, about $0.14^\circ\mathrm{C/min}$. If the coils can still hold the rack inlet temperatures within limits until the water has warmed by an illustrative 5 K, the loop provides $418\,\mathrm{MJ/K} \times 5\,\mathrm{K} = 2090\,\mathrm{MJ}$ of storage and $2090\,\mathrm{MJ}/1\,\mathrm{MW} = 2090\,\mathrm{s}$, about 35 minutes of thermal ride-through, enough to restart the chiller or bring a standby one online. Dedicated chilled-water storage tanks and other thermal storage amplify this buffer by adding mass that is kept cold for the purpose. The ride-through is real only if the water keeps moving: a loop whose pumps stop with the chiller stores the same energy but cannot deliver it to the coils.

The battery analogy. The energy balance of a battery and that of a thermal store have the same structure,

$$\frac{\mathrm{d}E_{\mathrm{battery}}}{\mathrm{d}t} = P_{\mathrm{in}} - P_{\mathrm{out}}, \qquad \frac{\mathrm{d}E_{\mathrm{thermal}}}{\mathrm{d}t} = \dot{Q}_{\mathrm{in}} - \dot{Q}_{\mathrm{out}}, \tag{8.14}$$

and the thermal store has a kind of thermal state of charge, normally expressed through temperature rather than energy: the colder the loop relative to its allowed upper temperature, the more heat it can still absorb. The 2090 MJ of the example equal $2090/3600 \approx 0.58\,\mathrm{MWh}$, the capacity of a sizable battery, held as the capacity of cold water to absorb heat. As with a battery, the useful content is bounded by limits on both sides, here the temperature at which the coils stop holding the racks within limits.

8.4 Propagation, delays and diagnosis

A disturbance does not affect every temperature at once. This section follows a load change through the facility, shows how the order and timing of the responses identify the cause of an excursion, and separates the three kinds of lag a record can contain: thermal inertia, transport delay and sensor dynamics.

8.4.1 Propagation through subsystems

An increase of IT load travels through the facility as a chain of effects,

$$P_{\mathrm{IT}} \uparrow \rightarrow \dot{Q}_{\mathrm{rack}} \uparrow \rightarrow T_{\mathrm{hot}} \uparrow \text{ (rack exhaust)} \rightarrow T_{\mathrm{return}} \uparrow \text{ (return air)} \rightarrow T_{\mathrm{CHW,r}} \uparrow \text{ (chilled-water return)} \rightarrow \text{chiller load} \uparrow,$$

and the links do not respond simultaneously. Each subsystem adds a capacitance that has to be charged and often a distance that the air or water has to travel: the rack heat appears in the exhaust within a minute or so, the return air at the cooling unit mixes the exhaust of many racks and follows more slowly, and the chilled-water return responds only after the coil has warmed and the water has carried the change back to the plant.

Figure 8.4 illustrates the chain for a load step at $t = 0$. The horizontal axis is minutes; the vertical axis is the fraction of each variable's final change completed, so that every variable runs from 0 to 1 and their timing can be compared directly. The gray step is the IT power. The three curves are the rack exhaust, the return air and the chilled-water return, each modeled as an illustrative delay followed by a first-order lag; the dotted line marks 63 % of the change, which the three reach at about 1, 3.5 and 9 minutes. Each subsystem shifts and smooths the change it passes on, so the plant sees a load step as a slow ramp several minutes late.

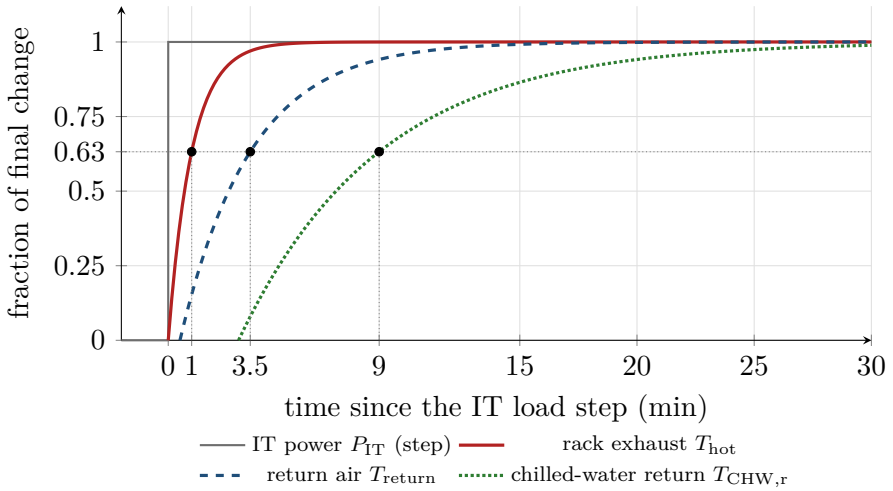


Figure 8.4: Propagation of an IT load step, illustrative. Each downstream variable is modeled as a delay t_d followed by a first-order lag τ, fraction $1 - e^{-(t-t_d)/\tau}$ for $t \geq t_d$: rack exhaust $t_d = 0$, $\tau = 1$ min; return air $t_d = 0.5$ min, $\tau = 3$ min; chilled-water return $t_d = 3$ min, $\tau = 6$ min. Each subsystem delays and smooths the change.

8.4.2 Dynamic signatures

Because each subsystem responds with its own lag, the order in which variables change carries information about where a disturbance began. If the rack inlet temperatures rise first and the chilled-water return follows minutes later, the disturbance entered on the air side, near the racks, and propagated toward the plant. If the chilled water changes first and the rack temperatures

follow, the fault lies upstream in the plant and propagated toward the racks. The diagnostic content is not only *what* changed but *when* each variable changed relative to the others; a fault leaves a dynamic signature, a pattern of onsets and lags, as well as a set of final values.

8.4.3 Three competing faults

Suppose the observed symptom is a rise of the rack inlet temperature $T_{\text{rack,in}}$. Three hypotheses compete to explain it.

- H_1, *an increase of IT load*: P_{IT} rises first, then the temperatures follow.
- H_2, *a degradation of a cooling-unit (CRAH) fan*: the air mass flow $\dot{m}_{\text{air}}$ falls first, then the rack temperatures rise, with little change in IT power.
- H_3, *a chiller problem*: the chilled-water supply temperature $T_{\text{CHW,s}}$ rises before the rack inlet temperatures do.

All three end with the same symptom. In a single snapshot taken after the rise they can be hard to separate, because each can produce the same inlet temperature. Their temporal ordering separates them.

Figure 8.5 draws the three signatures as timing diagrams. Each column is one hypothesis; each row is one measured signal, from top to bottom IT power, cooling-unit airflow, chilled-water supply temperature and rack inlet temperature, drawn schematically against time with the onset of the fault marked by the dashed vertical line. In each column the red trace is the signal that changes first, the blue trace the one that changes second, and the gray traces are signals that change little. The symptom row is blue in every column; the first-changing row is different in each, so the record of which signal moved first, and by how much it led the symptom, points to one hypothesis. Stated as inference, the time series raises the probability of the hypothesis whose predicted ordering matches and lowers the others, where a single value of $T_{\text{rack,in}}$ would leave them almost equal.

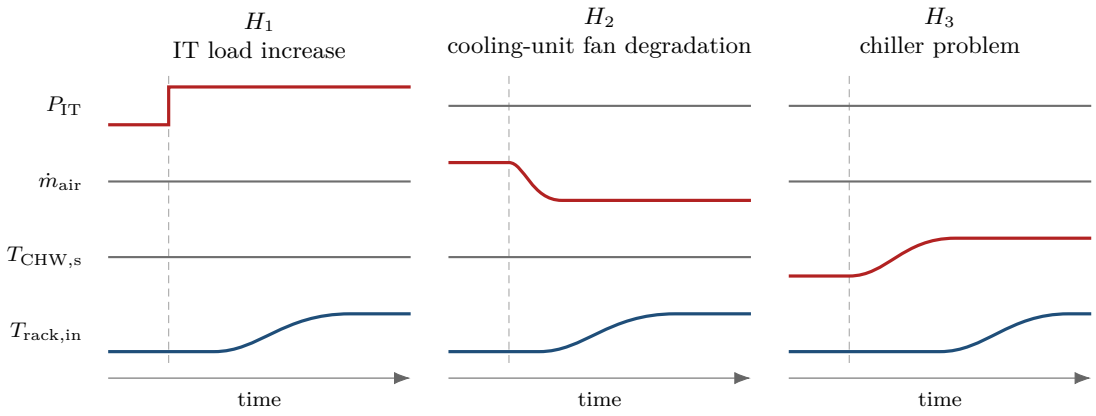


Figure 8.5: Timing signatures of three faults that end in the same symptom, a rise of the rack inlet temperature (schematic). Dashed line: onset of the fault. Red: first signal to change; blue: second (the symptom); gray: little change. H_1: IT power leads; H_2: airflow falls first with IT power flat; H_3: chilled-water supply temperature rises first and the rack inlet follows after the plant and transport lags.

Principle 8.2. *When several faults produce the same symptom, the order and timing in which the measured variables change, not their final values, is often what separates them.*

8.4.4 Dynamics against snapshots

A snapshot is a set of values at one instant; a dynamic record adds the rates and the order of change. Two hypotheses may both predict a rack temperature of $T = 28°\mathrm{C}$ at a given instant: one, say, a load that rose an hour ago, so that the rack is now at its new equilibrium with $\mathrm{d}T/\mathrm{d}t \approx 0$; the other a cooling degradation that began minutes ago, so that the rack is passing through 28°C on its way up, with $\mathrm{d}T/\mathrm{d}t > 0$. The snapshot cannot distinguish them; the time series can.

To make it concrete with illustrative numbers, let the second hypothesis predict $\mathrm{d}T/\mathrm{d}t = +0.3°\mathrm{C/min}$, and let the sensor have a random error with a standard deviation of 0.1 K. Over the next five minutes the first hypothesis predicts readings near 28.0°C throughout; the second predicts readings rising to about 29.5°C. The predictions differ by 1.5 K, fifteen times the noise, so five minutes of record leaves essentially no doubt about which hypothesis holds, while the single reading at 28°C fits both equally. Dynamics diagnose better than snapshots because the rate of change is a second, independent consequence of each hypothesis, and in a system with storage the rate is often the more informative one.

8.4.5 Delay versus time constant

Two kinds of lag must be kept apart. A *time constant* produces a gradual response that begins at once: after a step, the output starts moving immediately and approaches its final value as $1 - e^{-t/\tau}$. A *pure delay* produces no response at all until a time t_d has passed, after which the output reproduces the input, shifted:

$$T(t) = f\big(P(t - t_d)\big). \tag{8.15}$$

Data centers show both, often in series. Thermal inertia gives time constants; fluid transport gives delays, because a temperature change created at one point in an air or water stream reaches a distant sensor only after the fluid has traveled there.

Figure 8.6 compares the two on one axis. The horizontal axis is seconds since a step of the input; the vertical axis is the fraction of the final change seen at the output. The gray step is the input; the red trace is a pure delay of 100 s, which shows nothing for 100 s and then the whole step; the blue curve is a first-order lag with a time constant of 100 s, which starts at once and reaches 63 % at 100 s; the green dashed curve is the two in series, a delay followed by a lag. At $t = 100\,\mathrm{s}$ the three outputs read 0 (or 1 an instant later), 0.63 and 0. The difference shows in the first samples after a change: a lag moves at once with a finite slope, a delay does not move at all.

Transport delay in water. A temperature change at the chiller travels with the water. For an illustrative run of $L = 200\,\mathrm{m}$ of piping at a velocity of $v = 2\,\mathrm{m/s}$, the transport delay to a remote cooling unit is

$$t_d = \frac{L}{v} = \frac{200\,\mathrm{m}}{2\,\mathrm{m/s}} = 100\,\mathrm{s} \approx 1.7\,\mathrm{min}. \tag{8.16}$$

For 1.7 minutes after the chiller's supply temperature changes, the remote unit sees the old water. The delay varies inversely with flow: at half the velocity it doubles to 200 s. It matters when sensor records from different points are aligned: comparing the chiller's supply temperature with the coil inlet temperature at the same timestamp compares two different parcels of water, and the comparison is meaningful only after shifting one record by t_d, with t_d itself depending on the flow at that time.

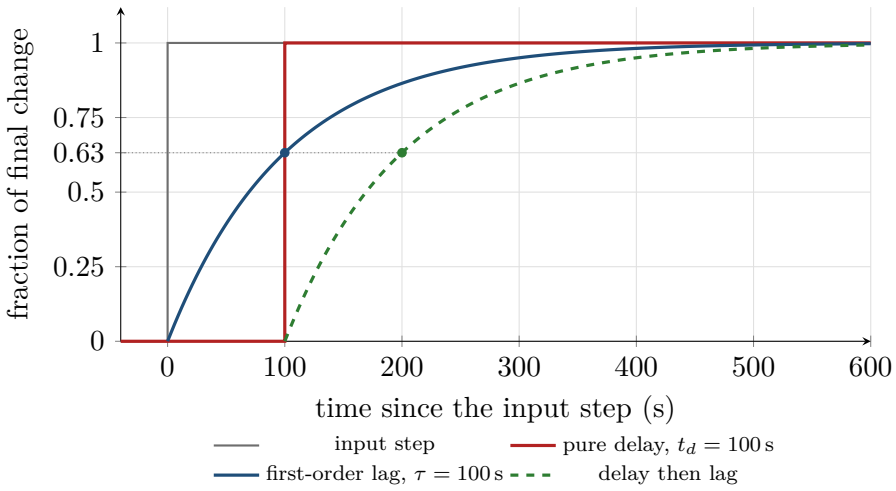


Figure 8.6: Pure delay against time constant. Input: a unit step at $t = 0$. Pure delay: $y(t) = 1$ for $t \geq t_d$, 0 before, with $t_d = 100\,\mathrm{s}$. First-order lag: $y(t) = 1 - e^{-t/\tau}$, $\tau = 100\,\mathrm{s}$. Delay then lag: $y(t) = 1 - e^{-(t-t_d)/\tau}$ for $t \geq t_d$. The values are illustrative.

8.4.6 Sensor delay and other sources of lag

A temperature sensor is itself a small thermal mass: its sensing element has to warm or cool to the temperature of the fluid around it through a thermal resistance. Its reading T_m therefore follows the true temperature T_{true} as a first-order system with its own time constant τ_s,

$$\tau_s \frac{\mathrm{d}T_m}{\mathrm{d}t} = T_{\mathrm{true}} - T_m. \tag{8.17}$$

The measured temperature lags the true one. For a true temperature rising at a constant rate r, the reading settles at a constant offset behind it: substituting $T_m = T_{\mathrm{true}} - \delta$ with $\mathrm{d}T_m/\mathrm{d}t = r$ gives $\delta = r\,\tau_s$. An illustrative sensor with $\tau_s = 30\,\mathrm{s}$ in air rising at $0.5^{\circ}\mathrm{C}$/min reads $0.5 \times 0.5 = 0.25^{\circ}\mathrm{C}$ low, and reports each value 30 s late. A sensor in a thermowell, or in still air, can have a time constant of minutes.

An observed lag in a temperature record can therefore come from four different sources, which must not be confused: Table 8.2 lists them with their form, their signature in a record, and how each is recognized. Thermal inertia belongs to the process and carries information about its capacitance; transport delay belongs to the piping and ductwork and depends on flow; sensor dynamics belong to the instrument; and the sampling and logging of the software add a delay of up to one polling interval and a staircase shape. Attributing a sensor lag to thermal inertia overestimates the capacitance; attributing a logging delay to transport misplaces a fault. Chapter 9 treats the instruments themselves.

8.5 Control interaction and dynamic capacity

Lags and delays shape how control loops behave and how much overload a facility can tolerate for how long. This section describes the oscillation that delays induce in control loops, the interaction of several loops, the difference between steady and dynamic capacity, and the time to limit as the operational form of thermal risk.

8.5.1 Control loops and oscillation

Consider a controller that holds a cold-aisle temperature by adjusting the fan speed of a cooling unit. The temperature rises; the controller raises the fan speed; the airflow rises at once, but

Table 8.2: Sources of lag in a temperature record. Values are illustrative.

Source	Form	Signature in a record	How it is recognized
Thermal inertia	$R_{\text{th}}C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = T_\infty - T$	response starts at once, exponential approach; seconds to hours	scales with mass; same for every sensor on the volume
Fluid transport	$T(t) = f(P(t - t_d))$, $t_d = L/v$	nothing until t_d, then the shape of the upstream change	varies inversely with flow; grows with distance
Sensor dynamics	$\tau_s\,\mathrm{d}T_m/\mathrm{d}t = T_{\text{true}} - T_m$	small extra lag and smoothing; offset $r\tau_s$ on ramps	fixed for a given sensor and mounting; differs between sensor types
Sampling and logging	reading held until the next poll; timestamp at logging	staircase; delay of up to one polling interval	matches the polling interval; seen in the timestamps

the temperature falls only after the lags of Section 8.4.5. During that interval the controller sees the temperature still high and keeps raising the fan speed. When the temperature finally responds, the fan is running faster than needed, the temperature falls below the setpoint, the controller backs off, and the temperature rises again. An aggressive controller acting on a delayed response overshoots and then oscillates.

The usual controller is the proportional-integral (PI) law,

$$u(t) = K_p\,e(t) + K_i \int_0^t e(s)\,\mathrm{d}s, \qquad e(t) = T_{\text{set}} - T(t), \tag{8.18}$$

where u is the control input, here the fan speed command, e is the error between the setpoint and the measured temperature, K_p is the proportional gain and K_i the integral gain. With the error defined this way, a temperature above the setpoint gives a negative error, so a cooling actuator, which must speed up when the temperature is high, uses negative gains (equivalently, positive gains on $T - T_{\text{set}}$). The proportional term reacts to the present error; the integral term accumulates past error and removes any steady offset. During a delay the integral keeps accumulating error that the actuator has already acted on, which is the mechanism of overshoot.

Poorly tuned gains have four kinds of consequence: oscillations, when the gains are too high for the lag of the loop; a slow response, when they are too low; unnecessary fan and pump energy, because an oscillating actuator spends energy pushing the temperature past the setpoint and back; and unstable interactions with neighboring loops. The larger the ratio of the loop's delay to its time constant, the more cautious the gains must be. Chapter 10 develops the control loops of a facility in detail.

8.5.2 Controllers that fight

A facility runs many loops at once, each on its own variable: cooling-unit fan control on air temperature or pressure, chilled-water valve control on the coil leaving-air temperature, pump control on differential pressure in the water loop, and chiller control on the chilled-water supply temperature. Each is local: it observes its own variable and acts on its own actuator. Because the variables are coupled through the heat and mass balances, one loop's action is another loop's disturbance. The fan loop raises the airflow to cool a warm aisle. With the IT heat unchanged, more air means a smaller temperature rise across the racks, so the return air entering the coil is cooler, and the air leaving the coil falls below the setpoint of the valve loop. The valve loop reads this as overcooling and closes the chilled-water valve, reducing water flow; the coil then

cools each kilogram of air less, the supply air warms back up, the aisle does not improve, and the fan loop raises the airflow further. One controller raises airflow while another reduces water flow; the two push against each other, each correct by its own measurement, and the fans run fast against a throttled coil. A local control action is not the same as a global improvement of the system, and when loops with similar time scales share a coupled process, their combined behavior has to be designed, not assumed.

8.5.3 Dynamic and steady-state capacity

Steady-state capacity asks whether the cooling can remove the heat indefinitely. Dynamic capacity asks whether the temperatures stay within limits for the duration of a particular event. The two answers differ because of storage. Take an illustrative facility with 10 MW of cooling capacity whose IT load suddenly rises to 12 MW. In steady state the situation is unsustainable: the cooling falls 2 MW short, and no equilibrium within limits exists. Yet the overload may be safe for five minutes, because the stored thermal capacity delays the temperature rise.

With an illustrative effective capacitance of 600 MJ/K for the parts of the facility that absorb the excess, equation (8.1) gives a rise of $2 \times 10^6 / 6 \times 10^8 = 0.0033\,\mathrm{K/s}$, or $0.2^\circ\mathrm{C/min}$. Five minutes of overload add $2\,\mathrm{MW} \times 300\,\mathrm{s} = 600\,\mathrm{MJ}$ of heat and raise the temperature by 1 K. In general, an overload $\Delta\dot{Q}$ held for a time D is tolerable while

$$\Delta\dot{Q}\, D \;\le\; C_{\mathrm{th}}\,(T_{\max} - T_0), \tag{8.19}$$

that is, while the excess heat does not exceed the heat the thermal mass can absorb within the temperature margin. With a 5 K margin the example tolerates $D = 600 \times 5/2 = 1500\,\mathrm{s}$, 25 minutes; an overload of 4 MW would be tolerable for half that. Both the magnitude and the duration of an overload matter, and a capacity statement that gives only the magnitude is incomplete. Equation (8.19) is the saturated-cooling case of Section 8.1.4; when the cooling still has a temperature-dependent part it errs on the safe side.

Principle 8.3. *Thermal capacity has a duration as well as a magnitude: an overload that no steady state can sustain may be tolerable for a time set by the stored capacity and the temperature margin, $D \le C_{\mathrm{th}}(T_{\max} - T_0)/\Delta\dot{Q}$.*

8.5.4 Thermal risk over a horizon

A static statement of thermal risk is the probability that the required cooling exceeds the available cooling, $\Pr(\dot{Q}_{\mathrm{IT}} > \dot{Q}_{\mathrm{avail}})$. It says whether the facility is short of cooling, not what the shortage will do or when. Because temperatures respond with storage and lag, the operational form of the risk has a time dimension: the probability that a rack temperature exceeds its limit at a given horizon h,

$$\Pr(T_{\mathrm{rack}}(t+h) > T_{\max} \mid \text{data up to } t), \qquad h = 1,\ 5,\ 10,\ 30 \text{ min}. \tag{8.20}$$

The two statements can differ sharply. A 2 MW shortfall with a large thermal mass can have a probability of one of being a shortfall and a probability near zero of producing a violation within five minutes; a rack with no shortfall at all but a large recirculation fraction can be close to its limit already. The horizon-indexed risk is the quantity an operator acts on: it says how much time remains, and with what confidence. Computing it requires the dynamic model of this chapter, the present state, and the uncertainty in both.

8.5.5 Time to limit

The simplest summary of the horizon-indexed risk is the time to limit. For a rack inlet at $T_0 = 25°\mathrm{C}$, a limit of $T_{\max} = 35°\mathrm{C}$ and a present rate of rise of $\mathrm{d}T/\mathrm{d}t = 0.5°\mathrm{C/min}$, a linear extrapolation gives

$$t_{\text{limit}} = \frac{T_{\max} - T_0}{\mathrm{d}T/\mathrm{d}t} = \frac{35 - 25}{0.5}\,\mathrm{min} = 20\,\mathrm{min}. \tag{8.21}$$

The statement that the limit is twenty minutes away is more useful to an operator than the statement that cooling capacity is insufficient: it says how urgent the response is and which responses are fast enough.

The linear extrapolation is conservative when the rise is first order. If the temperature approaches an equilibrium T_∞ with time constant τ, equation (8.8) reaches $T_{\max}$ only if $T_\infty > T_{\max}$, at

$$t_{\text{limit}} = \tau \ln \frac{T_\infty - T_0}{T_\infty - T_{\max}}, \tag{8.22}$$

which is always later than the linear estimate, because the rise slows as it goes. With $T_\infty = 45°\mathrm{C}$ and the same initial rate of $0.5°\mathrm{C/min}$, so that $\tau = (45 - 25)/0.5 = 40\,\mathrm{min}$, the limit is reached at $40 \ln(20/10) = 27.7\,\mathrm{min}$ rather than twenty; with $T_\infty \le 35°\mathrm{C}$ it is never reached. The linear form is the right default when the cooling is saturated or the equilibrium is unknown.

Uncertain time to limit. The present rate of rise is not known exactly. It depends on the rack power, the airflow, the effective capacitance and the sensor bias, each of which is uncertain, and it is computed from noisy, possibly lagged readings. The time to limit is therefore a distribution, $p(t_{\text{limit}} \mid y)$, given the measurements y, and the operational statement is a probability such as $\Pr(t_{\text{limit}} < 10\,\mathrm{min}) = 0.15$. As an illustrative model, let the rate be lognormal with median $0.5°\mathrm{C/min}$ and logarithmic standard deviation 0.669. Then $t_{\text{limit}} = 10\,\mathrm{K}/r$ is lognormal with median 20 min, and $t_{\text{limit}} < 10\,\mathrm{min}$ exactly when $r > 1.0°\mathrm{C/min}$, which happens with probability $1 - \Phi(\ln 2/0.669) = 1 - \Phi(1.036) = 0.15$, where Φ is the standard normal distribution function. The same model puts the 15th and 85th percentiles of the rate at 0.25 and $1.0°\mathrm{C/min}$, and those of the time to limit at 10 and 40 min.

Figure 8.7 draws the example. The horizontal axis is minutes from now; the vertical axis is the rack inlet temperature. The red dashed line is the limit $T_{\max} = 35°\mathrm{C}$. The solid line is the median trajectory at $0.5°\mathrm{C/min}$, and the shaded band spans the trajectories between the 15th and 85th percentiles of the rate. The time to limit is where a trajectory meets the limit: the median at 20 min, the band's edges at 10 and 40 min. A response that takes fifteen minutes to arrange is in time for the median trajectory but not for the fastest 15 % of plausible ones; the band, not the median line, decides whether to act now.

8.6 State space, estimation and identification

The models of the previous sections fit one general form, the state-space model, which is also the form that estimation and control methods take as input. This section writes it, identifies the states and parameters a record does not give directly, and asks what a record must contain for them to be determined.

8.6.1 The state-space form

Collect the temperatures that carry the facility's thermal memory in a state vector, for example

$$x = \left[T_{\text{rack}},\ T_{\text{return}},\ T_w\right]^{\mathsf{T}},$$

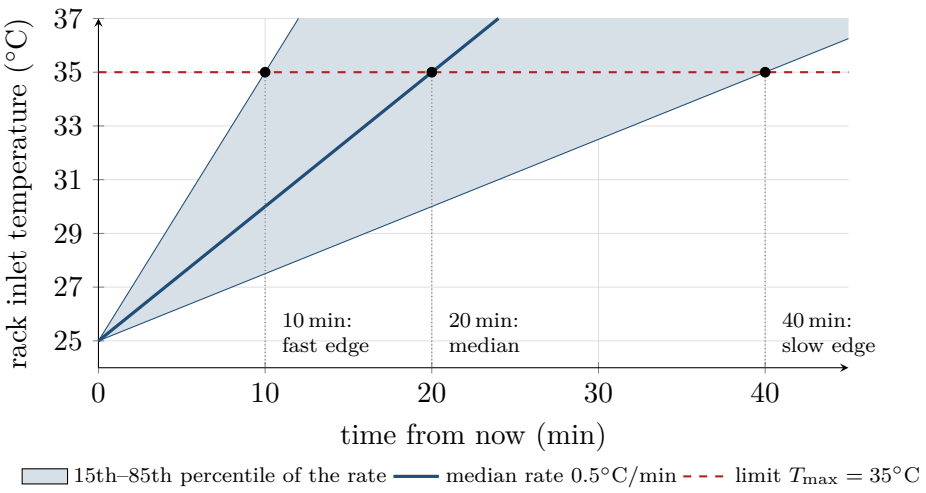


Figure 8.7: Time to limit as the intersection of a rising temperature with the limit. Illustrative: $T_0 = 25°\mathrm{C}$, $T_{\max} = 35°\mathrm{C}$, $T(t) = T_0 + r\,t$ with the rate r lognormal, median 0.5°C/min, logarithmic standard deviation 0.669; the band spans $r = 0.25$ to 1.0°C/min (15th to 85th percentile). $t_{\text{limit}} = 10\,\mathrm{K}/r$ has median 20 min and $\Pr(t_{\text{limit}} < 10\,\text{min}) = 0.15$.

the rack temperature, the return-air temperature and the chilled-water temperature. Their evolution is a set of coupled balances of the form of equations (8.11)–(8.12), written compactly as

$$\dot{x} = f(x, u, d, \theta), \tag{8.23}$$

with three kinds of input. The controls $u = [N_{\text{fan}},\ \dot{m}_w,\ T_{\text{CHW,set}}]^{\mathsf{T}}$ are what the operator or the control system sets: fan speed, water flow and the chilled-water supply setpoint. The disturbances $d = [P_{\text{IT}},\ T_{\text{amb}}]^{\mathsf{T}}$ act on the system without being chosen: IT power and outdoor temperature. The parameters $\theta = [R_1,\ R_2,\ C_1,\ C_2,\ \ldots]^{\mathsf{T}}$ are the resistances and capacitances, constant or slowly varying. The distinction matters because the three are handled differently: controls are decided, disturbances are forecast, and parameters are estimated.

Discrete time. Measurements arrive at discrete instants, so the model is used in its discrete form,

$$x_{k+1} = f_d(x_k, u_k, d_k, \theta), \qquad y_k = h(x_k) + \epsilon_k, \tag{8.24}$$

where f_d advances the state over one sampling interval, as the recursion of equation (8.9) does for one volume; h maps the state to what the sensors report; and ϵ_k is the measurement error. This pair of equations is the form required by state estimation, Kalman filtering and Bayesian filtering, which estimate x_k from the record $y_1, \ldots, y_k$; by forecasting, which runs f_d forward from the estimated state; and by model-based control, which chooses u_k to keep the forecast within limits. The time to limit of Section 8.5.5 is such a forecast, run until the predicted temperature crosses $T_{\max}$.

8.6.2 Hidden thermal states

Not every state is measured. The rack inlet temperature $T_{\text{rack,in}}$ is measured by a sensor; the chip temperature, the internal temperature of a rack and the effective thermal energy stored in its mass usually are not, or are available only through separate server telemetry. In the book's terms, the rack inlet is measured, the heat input $\dot{Q}_{\text{IT}} = P_{\text{IT}}$ is calculated from a power reading, and the chip temperature and the stored energy are inferred. Dynamic measurements make the inference possible: the way a measured temperature responds to a known change carries the

signature of the unmeasured masses behind it, as the slow component of Figure 8.3 carries that of the server air. A static record cannot do this; a hidden state that never changes leaves no trace in the measured ones.

8.6.3 Parameters from transients

A known step of the heat input is an experiment. After a step $\Delta\dot{Q}$, the first-order response of equation (8.8) is determined by its final change and its speed, and each carries one parameter:

$$R_{\mathrm{th}} = \frac{\Delta T_\infty}{\Delta \dot{Q}}, \qquad C_{\mathrm{th}} = \frac{\tau}{R_{\mathrm{th}}}. \tag{8.25}$$

The final rise informs R_{th}; the response time informs $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$, and hence C_{th}. For the rack of Section 8.2.3: a 2 kW step producing a final rise of 4 K gives $R_{\mathrm{th}} = 4/2000 = 0.002\,\mathrm{K/W}$; the 63 % point reached after ten minutes gives $\tau = 600\,\mathrm{s}$ and $C_{\mathrm{th}} = 600/0.002 = 300\,000\,\mathrm{J/K}$. Disturbances that happen anyway, a batch job that starts, a cooling unit that stops for maintenance, are informative experiments of this kind, provided the size and timing of the input change are recorded.

8.6.4 Identifiability of resistance and capacitance

Identification can still fail. With sparse measurements, many $(R_{\mathrm{th}}, C_{\mathrm{th}})$ pairs can reproduce the same measured trajectory. Equation (8.8) shows why: the trajectory depends on the parameters only through the time constant $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$ and the final change $R_{\mathrm{th}}\,\Delta\dot{Q}$. If the heat that actually reaches the measured volume is known, the final change gives R_{th} and the two separate. If it is not, because only part of the rack heat enters the measured air, or because the step size was not recorded, then the amplitude fixes only the product $R_{\mathrm{th}}\,\Delta\dot{Q}$, and perhaps only τ is identified: R_{th} and C_{th} cannot be separated. The same happens when the record covers only the first part of a transient, which fixes the initial slope $\Delta\dot{Q}/C_{\mathrm{th}}$, or only steady periods, which fix $R_{\mathrm{th}}\,\Delta\dot{Q}$.

Figure 8.8 shows the case of two parameter pairs with the same time constant. Both panels have minutes since a 2 kW step on the horizontal axis. The left panel shows the temperature rise in kelvin for pair A, $R_{\mathrm{th}} = 0.002\,\mathrm{K/W}$ and $C_{\mathrm{th}} = 300\,\mathrm{kJ/K}$, and pair B, $R_{\mathrm{th}} = 0.004\,\mathrm{K/W}$ and $C_{\mathrm{th}} = 150\,\mathrm{kJ/K}$: both have $\tau = 600\,\mathrm{s}$, but their equilibria differ, 4 against 8 K. The right panel shows the same two responses divided by their final values: they coincide. A record that carries only the normalized shape, because the heat input is uncertain, cannot tell A from B; one that carries a known input can. In the language of inference this is a property of the shape of the posterior: the probability given the data concentrates along the curve $R_{\mathrm{th}}C_{\mathrm{th}} = \tau$ in the $(R_{\mathrm{th}}, C_{\mathrm{th}})$ plane, a straight line $\ln R_{\mathrm{th}} + \ln C_{\mathrm{th}} = \ln \tau$ in logarithmic coordinates, rather than around a point, and the marginal uncertainty of each parameter is large even though τ is known precisely.

Dynamic data need dynamics. The information a record holds about a capacitance comes from transients. If P_{IT} never changes and the temperatures stay nearly constant, then $C_{\mathrm{th}}\,\mathrm{d}T/\mathrm{d}t \approx 0$ and the balance holds for any value of C_{th}: the data say little about it. The sensitivity of the step response to the capacitance makes this precise. Differentiating equation (8.8) with $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$,

$$\frac{\partial T}{\partial C_{\mathrm{th}}} = (T_0 - T_\infty)\,\frac{t}{R_{\mathrm{th}}C_{\mathrm{th}}^2}\,e^{-t/\tau} = \frac{T_0 - T_\infty}{C_{\mathrm{th}}}\,\frac{t}{\tau}\,e^{-t/\tau},$$

which is zero when there is no step ($T_0 = T_\infty$), zero at the instant of the step, and vanishes again once the response has settled. It is largest at $t = \tau$, where $(t/\tau)e^{-t/\tau}$ has its maximum e^{-1}. The samples that carry information about the capacitance are those taken around one time constant after a change; a record of steady operation, however long, carries almost none.

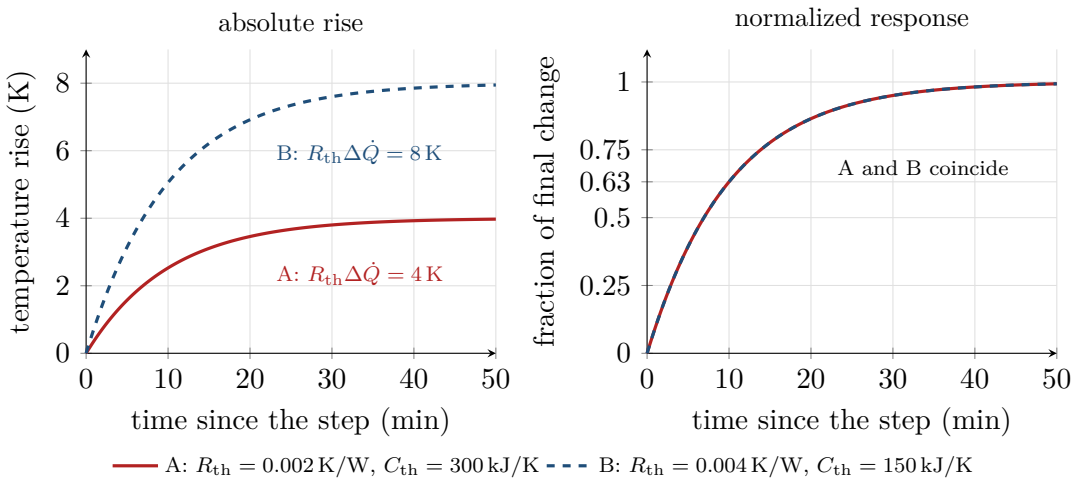


Figure 8.8: Two parameter pairs with the same time constant, $\tau = R_{\text{th}}C_{\text{th}} = 600\,\text{s}$, after an illustrative 2 kW step: $\Delta T(t) = R_{\text{th}}\,\Delta\dot{Q}\,(1 - e^{-t/\tau})$. Left: the equilibria differ (4 and 8 K). Right: the normalized responses $1 - e^{-t/\tau}$ are identical. Without a known heat input, only τ is identified.

8.6.5 Active testing

When natural disturbances are too rare or too poorly recorded, they can be created. A deliberate, small perturbation, a change of rack load with a synthetic workload or a load bank, or a change of a cooling unit's fan speed N_{fan}, followed by observation of the response, allows R_{th}, C_{th} and the airflow $\dot{m}_{\text{air}}$ to be inferred from equation (8.25) and its multi-mass extensions. This is system identification by active testing. The perturbation must be safe, small enough that the predicted rise, $R_{\text{th}}\,\Delta\dot{Q}$ for a load step, stays well inside the temperature margin, and informative, large enough that the response stands above the sensor noise and long enough, a few time constants, that both the speed and the final change are seen. For the rack of Section 8.6.3, a 0.5 kW step gives a 1 K rise, ten times an illustrative 0.1 K sensor noise, and thirty minutes of record cover three time constants.

8.6.6 Value of information

Suppose two thermal models are both plausible given the record: one in which the temperatures escalate fast after a loss of cooling, one in which they escalate slowly. The decisions they imply differ, so it is worth an experiment to separate them, and the best experiment is the one whose outcome is expected to separate their predictions most. A fan-speed change, a rack-load change and a valve change are all possible; their value differs.

Table 8.3 works an illustrative case. Model A has a time constant of 5 min and model B of 20 min, with the same resistance; the sensor noise is 0.1 K; each experiment is observed for five minutes. A 10 % fan-speed reduction that raises the equilibrium by 2 K under both models produces after five minutes $2(1 - e^{-1}) = 1.26\,\text{K}$ under A and $2(1 - e^{-0.25}) = 0.44\,\text{K}$ under B: a separation of 0.82 K, about eight noise standard deviations. A rack-load change of 0.25 kW with $R_{\text{th}} = 0.002\,\text{K/W}$ raises the equilibrium by only 0.5 K and separates the models by 0.21 K, about two standard deviations. A chilled-water valve change reaches the rack only after the transport and coil lags of Section 8.4.1, and within five minutes its predictions under the two models barely differ. The fan-speed test is the most informative; it is also the one whose risk must be checked against the margin, since it is the one that warms the racks most.

Table 8.3: Expected separation of two thermal models by three candidate experiments, illustrative. Models A and B differ in time constant (5 and 20 min); observation window 5 min; sensor noise 0.1 K.

Experiment	Equilibrium rise	Rise under A	Rise under B	Separation (noise units)
Fan speed $-10\,\%$	2 K	1.26 K	0.44 K	0.82 K (8.2)
Rack load +0.25 kW	0.5 K	0.32 K	0.11 K	0.21 K (2.1)
Chilled-water valve change	delayed	≈ 0	≈ 0	below 0.05 K (< 0.5)

8.6.7 The equations to keep and the dynamic state

Four equations carry the chapter:

$$C_{\mathrm{th}}\,\frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\mathrm{in}} - \dot{Q}_{\mathrm{out}}, \qquad \dot{Q} = \frac{T_1 - T_2}{R_{\mathrm{th}}}, \qquad \tau = R_{\mathrm{th}}C_{\mathrm{th}}, \qquad T(t) = T_\infty + (T_0 - T_\infty)\,e^{-t/\tau}. \tag{8.26}$$

The first is conservation with storage; the second is the transfer of heat between two temperatures T_1 and T_2 through a resistance; the third is the speed of a single volume; the fourth is its response to a step. Multi-mass models, delays, sensor lags and the state-space form are built from these pieces.

The dynamic state. A thermal system is described by more than its temperatures: it has temperature, stored thermal energy and a delayed response. That changes the operational question from whether the system is safe now to where, given its state and its dynamics, it is heading. Consider two racks with an illustrative limit of 35°C: rack A at 30°C and cooling at -0.2°C/min; rack B at 27°C and rising at 2°C/min. By temperature alone, A is the more concerning, 5 K from the limit against 8 K. By the dynamic state, A is moving away from its limit and has no time to limit at all, while B reaches 35°C in $(35 - 27)/2 = 4$ minutes. The pair $(T, \mathrm{d}T/\mathrm{d}t)$, the temperature and its rate of change, can matter more than temperature alone, and a monitoring system that ranks racks by temperature only ranks them in the wrong order here.

8.7 What to retain

Definition 8.1 (Thermal dynamics)**.** A data center's thermal state is a set of temperatures together with the energy stored behind them and the delays that separate cause from effect, governed by $C_{\mathrm{th}}\,\mathrm{d}T/\mathrm{d}t = \dot{Q}_{\mathrm{in}} - \dot{Q}_{\mathrm{out}}$ with time constants $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$; the operational question is where the temperatures are heading and how long until a limit is reached, not only whether they are within limits now.

Table 8.4 lists the chapter's objects and the subsection that defines each.

8.8 Exercises

Exercise 8.1. A hall dissipates $\dot{Q}_{\mathrm{IT}} = 1.5$ MW. After a cooling-unit failure the remaining cooling, at its maximum, removes 1.2 MW. The effective thermal capacitance is 120 MJ/K. (a) Compute the rate of temperature rise in °C/min, the rise after 15 minutes, and the time until the temperature has risen by 5 K. (b) Repeat the time to a 5 K rise if only the room air, 15 MJ/K, took part. (c) What does the comparison say about which masses must be coupled to the air for the larger capacitance to count?

Table 8.4: The objects of thermal dynamics and where each is defined.

Object	Expression	Role	Defined in
Dynamic heat balance	equation (8.1)	conservation with storage	8.1.1
Thermal capacitance	$C_{\text{th}} = m\,c_p$	energy per kelvin; sets the slope	8.1.2
Thermal resistance	$\dot{Q} = \Delta T / R_{\text{th}}$	heat transfer along a path	8.2.1
First-order model	equation (8.4)	loss grows with temperature	8.2.1
Equilibrium	$T_{\text{eq}} = T_{\text{cool}} + R_{\text{th}}\dot{Q}_{\text{IT}}$	where the temperature settles	8.2.2
Time constant	$\tau = R_{\text{th}}C_{\text{th}}$	63 % of a step at 1τ	8.2.3
Step response	equation (8.8)	continuous, asymptotic approach	8.2.4
Distributed lag	equation (8.10)	memory as geometric weights	8.3.2
Two-mass model	equations (8.11)–(8.12)	fast and slow components	8.3.3
Chilled-water store	$C_w = m_w c_{p,w}$	ride-through buffer	8.3.5
Dynamic signature	Figure 8.5	order of change locates a fault	8.4.3
Transport delay	$t_d = L/v$	nothing until t_d	8.4.5
Sensor lag	equation (8.17)	reading lags by τ_s	8.4.6
PI controller	equation (8.18)	overshoot under delay	8.5.1
Dynamic capacity	$\Delta\dot{Q}\,D \le C_{\text{th}}(T_{\max} - T_0)$	overload magnitude and duration	8.5.3
Horizon risk	equation (8.20)	risk with a time dimension	8.5.4
Time to limit	equations (8.21), (8.22)	time remaining, a distribution	8.5.5
State-space model	equations (8.23), (8.24)	form for estimation and control	8.6.1
Identifiable combination	$\tau = R_{\text{th}}C_{\text{th}}$	what an uncertain input leaves	8.6.4
Dynamic state	$(T, \mathrm{d}T/\mathrm{d}t)$	where the system is heading	8.6.7

Exercise 8.2. A rack zone has $R_{\text{th}} = 0.0015\,\text{K/W}$, $C_{\text{th}} = 400\,\text{kJ/K}$ and a cooling medium at $T_{\text{cool}} = 18°\text{C}$. (a) Compute the time constant and the equilibrium temperature at $\dot{Q}_{\text{IT}} = 4\,\text{kW}$. (b) The load steps from 4 to 6 kW at $t = 0$ with the zone at its old equilibrium. Compute the new equilibrium and the temperature after 10 and 20 minutes. (c) How long does the zone take to reach 26.5°C, and when is it effectively settled?

Exercise 8.3. A rack inlet reads $T_0 = 26°\text{C}$ against a limit $T_{\max} = 32°\text{C}$ and is rising at 0.4°C/min. (a) Compute the linear time to limit. (b) If the rise is first order toward $T_\infty = 36°\text{C}$ with the same initial rate, compute the time constant and the time to limit from equation (8.22), and say why it is later. (c) The rate is uncertain, lognormal with median 0.4°C/min and logarithmic standard deviation 0.5. Using the linear model, compute the median time to limit and $\Pr(t_{\text{limit}} < 10\,\text{min})$.

Exercise 8.4. A 3 kW synthetic workload is switched on in a rack. The rack inlet rises by a final 4.5 K and reaches 63 % of that rise after 8 minutes. (a) Compute R_{th}, τ and C_{th}. (b) It later emerges that the heat reaching the measured air may have been anywhere between 70 and 130 % of the 3 kW. Compute the resulting ranges of R_{th} and C_{th}, and state what remains exactly determined. (c) At what time after the step is a single reading most informative about C_{th}? (d) The chilled water feeding the rack's cooling unit travels 300 m from the plant at 1.5 m/s; by how much must the plant's supply-temperature record be shifted to compare it with the unit's coil-inlet record?

8.9 Connection to the TERRA framework

Thermal dynamics adds storage and time to the conservation graph of a data center; every object of the chapter has a place on it.

- **Thermal nodes carry storage.** Each control volume, a rack, an aisle, a return plenum, the chilled-water loop, is a node of the thermal-energy network whose balance is equation (8.1); its stored quantity is $C_{\text{th}}T$, and the steady balance of Chapter 5 is the same node with $\mathrm{d}T/\mathrm{d}t = 0$. The chilled-water loop is a node with a large C_w, which is what makes ride-through computable.

- **Edges carry heat, with resistance and delay.** The flow between two thermal nodes is $(T_1 - T_2)/R_{\text{th}}$ (equation (8.26)); the air and water edges that carry heat by moving mass add a transport delay $t_d = L/v$ that depends on the flow on the same edge. The two-mass model of equations (8.11)–(8.12) is two nodes and one shared edge.

- **Forward simulation gives risk with a horizon.** The state-space form of equation (8.24) is the graph stepped forward in time. Run from the current posterior over the states and parameters, it gives $\Pr(T_{\text{rack}}(t+h) > T_{\max})$ for each horizon and the distribution of the time to limit, as in Figure 8.7.

- **Hidden states and parameters are inferred, with their shape.** Chip temperatures, stored energy, R_{th} and C_{th} are unmeasured quantities given posteriors from transients. When the heat input to a node is uncertain, the posterior is a ridge along $R_{\text{th}}C_{\text{th}} = \tau$ (Figure 8.8); a prior on C_{th} from the mass inventory, equation (8.2), or a measured input separates them.

- **Faults are hypotheses scored on time series.** H_1 to H_3 of Section 8.4.3 are alternative perturbations of the graph, an input change, an edge flow change, a source temperature change, and the likelihood of a whole record, with its onsets and lags (Figure 8.5), ranks them where a snapshot leaves them tied.

- **The next experiment is chosen for information.** An active test is a deliberate input to the graph; the candidate with the largest expected separation of the competing models' predictions, as in Table 8.3, is the one to run, subject to the predicted rise staying inside the margin.

In summary, a data center's temperatures are states of coupled heat balances with storage, linked by resistances and transport delays; the response to a change is gradual and ordered in time, the order locates faults, the stored energy turns a steady-state capacity statement into one with a duration, and the useful operational quantity is the probability of reaching a limit within a horizon, which depends on parameters that only transients reveal. The next chapter turns to the sensors that produce these time series: what they measure, where they sit, and how their errors and delays enter the record.

Chapter 9

Sensors and instrumentation: what is measured, what is estimated and what stays hidden

9.1 The measurement equation

The previous chapters described a data center as a physical system: electrical power flowing through a distribution hierarchy, heat produced by the IT equipment, air and water carrying that heat away, and a thermal state that changes in time. This chapter turns to what the facility knows about that system. The question is

which physical quantities in a data center are directly measured, which can only be estimated from other measurements, and which important states remain hidden however the data are processed.

The answer matters because every diagnostic or inference system is constrained by the measurement architecture it is given. An algorithm can combine, filter and interpret measurements; it cannot create information that the sensors did not collect. The chapter follows the chain

$$\text{physical system} \rightarrow \text{sensors} \rightarrow \text{telemetry} \rightarrow \text{state estimates and diagnosis},$$

from the quantities sensors report, through the ways those reports err, to the checks that expose the errors, the structural limits that no check can remove, and the layers of estimation that turn raw readings into a statement of risk.

Jargon: telemetry, BMS, DCIM, bias, drift, residual, observability, virtual sensor, soft sensor, sensor fusion, representativeness

Telemetry is the automatic measurement of quantities at remote equipment and their transmission, as time-stamped records, to a central system. A *building management system* (BMS) is the supervisory software that monitors and controls the facility's mechanical and electrical plant: air handlers, pumps, chillers, valves and their alarms. *Data center infrastructure management* (DCIM) is software that tracks the infrastructure specific to a data center, such as rack power, capacity, floor space, IT assets and power paths. A sensor's *bias* is its systematic error, the nonzero average difference between its reading and the true value; *drift* is a slow change of that bias over time. A *residual* is the difference between a measured value and the value a model predicts, or between two calculations that should agree. *Observability* is the property that the available measurements contain enough information to determine a given state. A *virtual sensor* is a quantity calculated from measurements through a physical relation and reported as if it were measured. A *soft sensor* is an estimate produced by a model that combines measurements with empirical or learned relations and physical constraints. *Sensor fusion* is the combination of several measurements of different kinds into one estimate of a quantity, each weighted by its reliability. *Representativeness* is the degree to which a measurement taken at one point stands for the quantity it is used to represent, such as the inlet temperature of a whole rack.

Key idea

A sensor reports a function of the physical state plus an error, never the state itself. What a data center can know about itself is set by which functions are measured, how often and how accurately, and how they err; some important states, such as rack airflow, remain hidden however many temperature sensors are installed, and the honest output of an analysis keeps measured, inferable and not inferable quantities apart.

9.1.1 From physical system to telemetry

The physical system has states: temperatures at every point of the air and water, mass flows in every duct and pipe, power on every conductor, positions of every valve and breaker, and parameters such as the conductance of a coil or the resistance of a filter. Sensors are placed at a small number of points in that system. Each sensor converts a local physical effect, the resistance of a thermistor, the voltage of a current transformer, the deflection of a pressure diaphragm, into an electrical signal; a controller or meter digitizes the signal, attaches a timestamp and transmits it; and a supervisory system stores the result as telemetry. Only after that chain does anyone estimate a state or diagnose a fault.

Each link of the chain removes or distorts information. The sensor sees one point, not the field around it. The digitizer rounds the value and samples it at intervals. The clock that stamps it may be wrong. The network may drop or duplicate records. The label that names the point may be attached to the wrong equipment. A state estimate built on the telemetry inherits all of these effects, and the rest of the chapter treats them one at a time.

9.1.2 The measurement equation

The common form of every sensor is the *measurement equation*

$$y = h(x, \theta) + \epsilon, \tag{9.1}$$

where x is the state of the physical system, θ is a vector of parameters, h is the function the sensor actually responds to, y is the reported value and ϵ is the measurement error. The

equation says that a sensor reveals a function of the state, not the state. Three examples show the range of h:

$$\text{temperature:} \quad y_T = T + \epsilon_T, \tag{9.2}$$

$$\text{electrical power:} \quad y_P = VI\cos\varphi + \epsilon_P, \tag{9.3}$$

$$\text{pressure difference:} \quad y_{\Delta p} = p_1 - p_2 + \epsilon_p. \tag{9.4}$$

For a temperature sensor h is the identity at the sensor's location: the reading is the local temperature plus an error. For a power meter h is a product of three quantities, voltage V, current I and the cosine of the phase angle φ between them; the meter measures the first two and computes the product internally, so even a "direct" power reading is a calculation inside the instrument. For a pressure sensor h is a difference between two locations; the absolute pressures p_1 and p_2 are not reported.

Equation (9.1) separates three questions that the rest of the chapter keeps distinct. The first is *what* h is: which function of the state a sensor sees, and at which location. The second is *how* ϵ behaves: whether it is small, whether it averages to zero, whether it changes over time. The third is whether the collection of all the functions h carries enough information to determine the states and parameters of interest. A facility can answer the first two well and still fail the third.

9.2 What is measured

This section surveys the quantities a data center measures: electrical quantities and their metering hierarchy, temperatures, humidity, pressure differences, flows, discrete equipment states, server telemetry, and the two supervisory systems that collect them. For each, it states what the measurement gives and what it does not.

9.2.1 Electrical measurements

Electrical meters report voltage V, current I, real power P, reactive power Q (the part of the power that oscillates between source and load without doing work, from the inductance and capacitance of the equipment), apparent power $S = VI$, the power factor $P/S = \cos\varphi$, frequency f, and energy E, the time integral of real power. In a three-phase circuit with line-to-line voltage V, the real power is $P = \sqrt{3}\,VI\cos\varphi$. A rack power distribution unit on an illustrative 400 V three-phase feed carrying 32 A per phase at a power factor of 0.95 therefore draws $\sqrt{3} \times 400 \times 32 \times 0.95 \approx 21.1$ kW.

Meters sit at many levels of the power path described in Chapter 3: the utility service entrance, the medium- and low-voltage switchgear, the transformers, the uninterruptible power supplies (UPS), the power distribution units (PDUs), the busways, the rack PDUs and, sometimes, the servers themselves, whose power supplies report their own input power. Together they form a hierarchy of power measurements,

$$P_{\text{facility}}, \quad P_{\text{UPS}}, \quad P_{\text{PDU}}, \quad P_{\text{rack}}, \quad P_{\text{server}}, \tag{9.5}$$

each level the sum of the level below plus the losses and loads that branch off between them. The upper levels usually carry revenue-grade or switchgear meters of high accuracy; the lowest level, the server's own report, comes from instrumentation built for power management rather than for metering and is less accurate. Electrical power is the best measured quantity in a data center: it is metered at most branch points, often at one-second resolution, and it is the input from which the heat load of the IT equipment is calculated.

9.2.2 Hierarchical metering and the facility balance

The hierarchy makes a consistency check possible. Take an illustrative facility whose utility meter reads $P_{\text{facility}} = 12\,\text{MW}$ while the PDU meters sum to $\sum_i P_{\text{PDU},i} = 10.7\,\text{MW}$. The difference,

$$12 - 10.7 = 1.3\,\text{MW},$$

about 11 % of the facility power, is everything that is not delivered to the IT equipment: cooling (chillers, air handlers, fans), the losses of transformers and UPS, pumps, lighting, and auxiliary loads such as offices, security and controls. Without further meters the 1.3 MW is one unexplained number. If those branches are metered as well, the facility balance

$$P_{\text{facility}} \approx P_{\text{IT}} + P_{\text{cool}} + P_{\text{loss}} + P_{\text{other}} \tag{9.6}$$

can be tested, where P_{IT} is taken as the PDU sum, P_{cool} includes pumps, P_{loss} is the conversion loss of transformers and UPS (input minus output of each), and P_{other} is lighting and auxiliaries. With illustrative sub-meter readings $P_{\text{cool}} = 0.95\,\text{MW}$, $P_{\text{loss}} = 0.25\,\text{MW}$ and $P_{\text{other}} = 0.07\,\text{MW}$, the residual of the balance is

$$r = 12 - (10.7 + 0.95 + 0.25 + 0.07) = 0.03\,\text{MW} = 30\,\text{kW},$$

0.25 % of the facility power. If each meter is accurate to about 0.5 %, the utility meter alone is uncertain by ±60 kW, so a 30 kW residual is consistent with correct meters. A residual of 300 kW would not be: it would indicate a missing branch, a meter error or a mislabeled point. The ratio $P_{\text{facility}}/P_{\text{IT}} = 12/10.7 \approx 1.12$ is the power usage effectiveness that Chapter 11 examines; the balance shows that its accuracy depends on every meter in equation (9.6).

Figure 9.1 draws the hierarchy. Each box is a metering point; arrows run in the direction of power flow, from the utility meter at the top down through the UPS and PDUs to the racks and servers, with the cooling plant and the lighting and auxiliaries branching off at the second level. The server box is dashed because its reading is the server's own report and is not always available. The box on the right writes the balance of equation (9.6) with the illustrative numbers. The figure shows why hierarchical metering is more than bookkeeping: every parent meter is a check on the sum of its children, so an error anywhere in the tree appears as a residual at the level above it.

9.2.3 Temperature measurements

Temperature is the most widely measured quantity in a data center. It is measured at the server inlet and exhaust (by sensors on the server board), at the rack inlet and exhaust, in the cold aisle and the hot aisle, at the room return, at the supply and return of each computer room air handler (CRAH), on the chilled-water supply and return, $T_{\text{CHW,s}}$ and $T_{\text{CHW,r}}$, on the condenser-water supply and return, and in the outdoor air, T_{amb}. Thermistors, resistance thermometers and thermocouples are inexpensive, small and easy to install, so a large hall can carry thousands of temperature points.

Figure 9.2 draws typical sensor locations on a vertical section through one cold aisle. The CRAH on the left pushes cooled air into the underfloor plenum; the air rises through a perforated tile into the cold aisle between two rows of racks, passes through the racks, and leaves as hot exhaust into the hot aisles, from which it returns under the ceiling to the top of the CRAH. Blue arrows are cool air, red arrows warm air, and the dashed red arrow is hot exhaust curling over the top of the right-hand rack into its inlet. The numbered markers are sensors; their shape gives the kind of quantity and the key on the right gives each location. Temperature points (circles) are everywhere. There is one humidity point, two pressure differences, a water flow meter and the rack's electrical power; there is no marker for the airflow through either

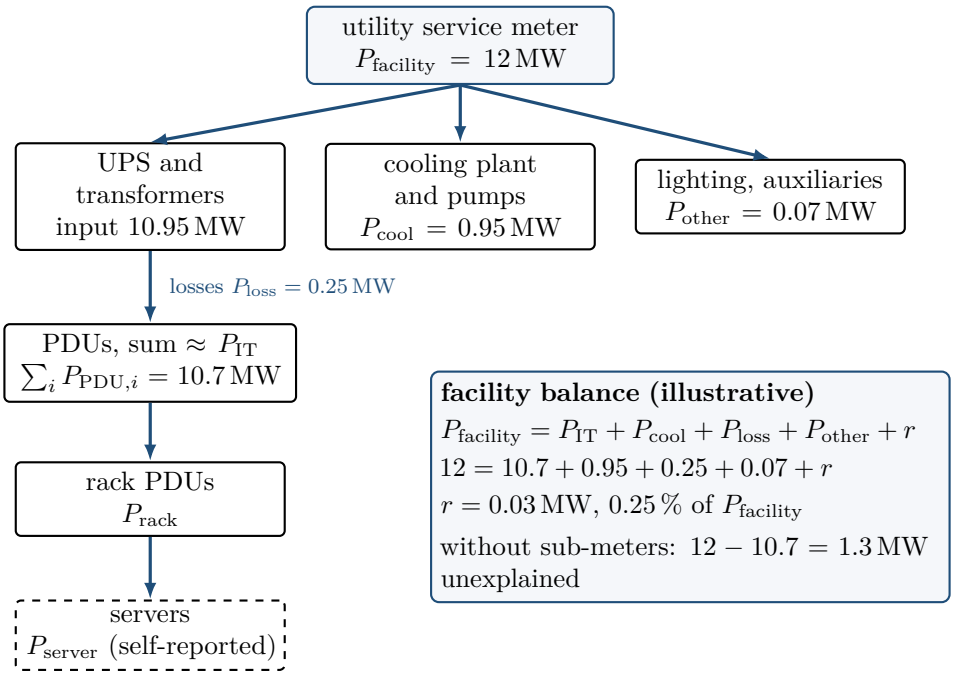


Figure 9.1: The power metering hierarchy from the utility service to the servers, with the facility balance of equation (9.6) (illustrative values). Each parent meter checks the sum of its children; the residual r is what the meters leave unexplained. The server level is dashed because it relies on the servers' own reports.

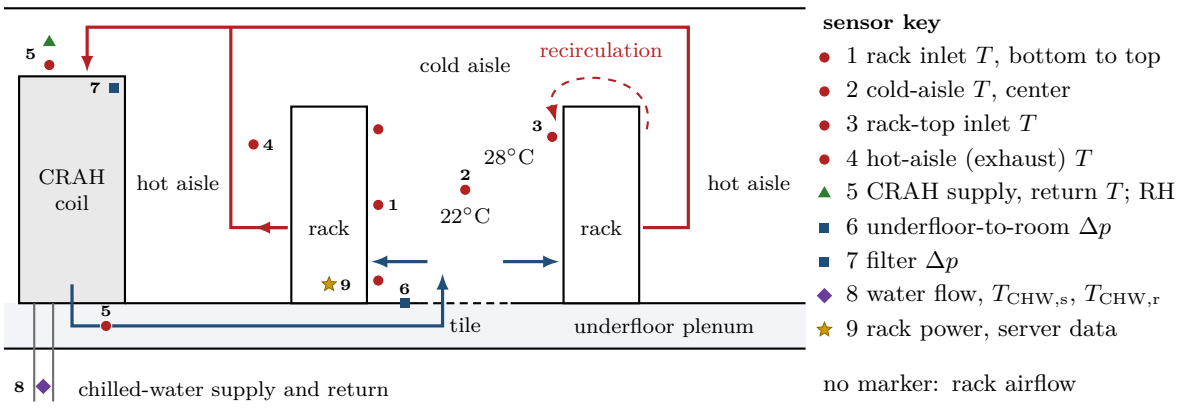


Figure 9.2: Typical sensor locations on a vertical section through a cold aisle (schematic, not to scale). Circles are temperatures, the triangle humidity, squares pressure differences, the diamond water flow and the star electrical power and server telemetry. Temperature points are plentiful; the airflow through each rack is not measured. The two temperatures are the illustrative values of Section 9.5.2.

rack, because that is rarely measured. The two labeled temperatures, 22°C in the middle of the cold aisle and 28°C at the top of the right-hand rack, are discussed in Section 9.5.2.

The ease of measuring temperature creates an imbalance in observability. Temperature is a result: it is set by the heat added and by the mass flow that carries the heat, through $\dot{Q} = \dot{m} c_p \Delta T$ of Chapter 5. Of the three quantities in that relation, heat is calculated from electrical power, temperature is measured densely, and mass flow, especially airflow, is measured rarely. A facility can therefore have a detailed map of the result and almost no direct measurement of one of its causes.

Many temperature sensors, still unobservable. Adding temperature sensors does not by itself remove that imbalance. A hall with 1000 temperature sensors may still leave unknown the airflow through each rack, the recirculation fraction α at each rack inlet (the share of inlet air that is hot exhaust, from Chapter 6), the heat-transfer coefficient UA of each coil, and slow hidden degradation such as a clogging filter. The reason is that many different combinations of these states produce similar temperature fields. A rack whose inlet is 3 K warmer than its neighbors may receive less supply air, may draw more recirculated air, or may sit next to a rack that has just increased its load; each combination moves the readings in nearly the same way. Adjacent sensors also move together, so the thousandth sensor adds little that the nine-hundred-ninety-ninth did not already say. A rough count makes the point: an illustrative hall of 200 racks has at least two unknown air states per rack, airflow and recirculation, or 400 unknowns, and if each of the 1000 sensors may carry its own bias, 1000 more. At any instant there are more unknowns than readings, and only structure, such as biases that change slowly while the air states change quickly, can separate them.

> **Principle 9.1.** *Many measurements are not the same as enough independent information. The number of states a set of sensors can determine is bounded by the number of independent ways the states move the readings, not by the number of sensors.*

9.2.4 Humidity and pressure

Humidity is reported as relative humidity RH, the ratio of the water vapor present to the maximum the air can hold at its temperature, and sometimes as dew point, the temperature at which that vapor would begin to condense. It matters for three reasons: condensation on cold coils and pipes, electrostatic discharge in very dry air, and the performance of evaporative cooling, which depends on how much more water the air can absorb. Relative humidity depends strongly on temperature: the same air, with the same moisture content, shows a lower RH when it is warmed and a higher RH when it is cooled, so the same parcel reads differently at the CRAH supply and at the rack exhaust. Moisture content (kilograms of water per kilogram of dry air) or dew point is conserved as air is heated or cooled without condensation, and is therefore the more informative quantity for tracking moisture through the air path.

Pressure is usually measured as a pressure difference Δp rather than an absolute value: between the underfloor plenum and the room, between the cold aisle and the hot aisle, across a filter, and across a pump or a coil in a water loop. Flow depends on pressure difference, so these readings are the closest the air side usually comes to a flow measurement. A pressure difference is not a flow, however. Across a fixed restriction the two are related by

$$\Delta p = K \, \dot{V}^2, \qquad \dot{V} = \sqrt{\Delta p / K}, \tag{9.7}$$

where K is a flow-resistance coefficient. With an illustrative tile coefficient $K = 200\,\mathrm{Pa\,s^2/m^6}$, a measured $\Delta p = 12.5\,\mathrm{Pa}$ gives $\dot{V} = \sqrt{12.5/200} = 0.25\,\mathrm{m^3/s}$. If cables under the floor partially block the tile and the true coefficient is $250\,\mathrm{Pa\,s^2/m^6}$, the same reading corresponds to

$\sqrt{12.5/250} = 0.224\,\mathrm{m^3/s}$, about 11 % less. Because the flow goes as the square root, a relative error in K produces half that relative error in the flow; but K is rarely known for each tile, filter or rack, and it changes as filters load and cables move. An uncertain K makes the flow uncertain however precise the pressure reading is.

9.2.5 Flow and why it matters

Flow is measured as a volumetric flow $\dot{V}$ or a mass flow $\dot{m}$: air flow $\dot{V}_{\mathrm{air}}$ or $\dot{m}_{\mathrm{air}}$, chilled-water flow $\dot{m}_{\mathrm{CHW}}$, and condenser-water flow $\dot{m}_{\mathrm{CW}}$. Water flow is commonly metered at major equipment, chillers, pumps and sometimes each CRAH coil, with ultrasonic or electromagnetic meters on the pipes. Air flow is measured in some ducts and air handlers, but the flow through an individual rack is rarely measured: an open rack has no duct with a straight run for an airflow station, and a flow hood held against a tile is a manual, occasional measurement. In practice rack airflow is estimated, not measured.

The absence matters because of the relation that ties heat, flow and temperature together,

$$\dot{Q} = \dot{m}\, c_p\, \Delta T. \tag{9.8}$$

Measuring T_{in} and T_{out} gives $\Delta T = T_{\mathrm{out}} - T_{\mathrm{in}}$, but without $\dot{m}$ the heat flow cannot be computed. A temperature rise of $\Delta T = 10\,\mathrm{K}$ in air, with $c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ/(kg\,K)}$, corresponds to

$$\dot{Q} = 1\,\mathrm{kg/s} \times 1.0 \times 10 = 10\,\mathrm{kW} \qquad \text{or} \qquad \dot{Q} = 100\,\mathrm{kg/s} \times 1.0 \times 10 = 1000\,\mathrm{kW} = 1\,\mathrm{MW},$$

depending on the mass flow; at an air density of about $1.2\,\mathrm{kg/m^3}$ these are 0.83 and $83\,\mathrm{m^3/s}$. The same 10 K is a single server or a whole hall. On the water side the same statement holds with $c_{p,w} = 4.18\,\mathrm{kJ/(kg\,K)}$: 1 MW at a 10 K rise needs $1000/(4.18 \times 10) \approx 23.9\,\mathrm{kg/s}$ of water. A temperature difference is half of a heat measurement; the flow is the other half.

9.2.6 Equipment states, commanded and actual

Much telemetry is discrete: ON or OFF, OPEN or CLOSED, AVAILABLE or FAILED, AUTO or MANUAL. Typical points are the position of each breaker, the state of each automatic transfer switch, whether each pump is running, whether each chiller is available, the state of each generator, the status of each valve, and the command sent to each fan. These states determine the topology of the system: which UPS feeds which PDU, which source a transfer switch has selected, which coils are in the chilled-water loop. A wrong status point is therefore not a small numerical error; it is a wrong graph, and every balance written on that graph is wrong with it.

Many actuators report three different quantities. The *commanded* state u_{cmd} is what the controller asked for; the *feedback* u_{feedback} is what the actuator reports, usually from a position sensor on its motor; and the *actual* state u_{actual} is the true position of the valve stem, damper blade or fan. In normal operation the three agree. A mechanical fault separates them: a valve commanded to $u_{\mathrm{cmd}} = 80\,\%$ open may be physically at $u_{\mathrm{actual}} = 20\,\%$ because the coupling between the actuator and the stem has slipped, or because the stem is seized. Whether the feedback reveals the fault depends on where its sensor is. If it measures the motor, it reports 80 % with the command, and the fault is invisible in the valve telemetry; only the consequence, a coil that removes less heat than its valve position implies, can expose it. This is a hidden-state problem: the true actuator position is a state to be inferred, not read.

Figure 9.3 shows such a fault. The upper part is a block diagram of the actuation chain: the controller sends the command to the actuator, whose motor position sensor produces the feedback; a linkage connects the actuator to the valve stem, and the stem position sets the water flow. The cross marks the slipped linkage. The three strips below plot the three signals against time in minutes, on a common vertical scale from 0 to 100 % open. At minute 5 the

command steps from 20 to 80 %; the feedback follows over the actuator's 90-second stroke; the actual position stays at 20 %. The telemetry agrees with itself, and the 60-point gap between what the system reports and what the valve is doing appears in no reported signal.

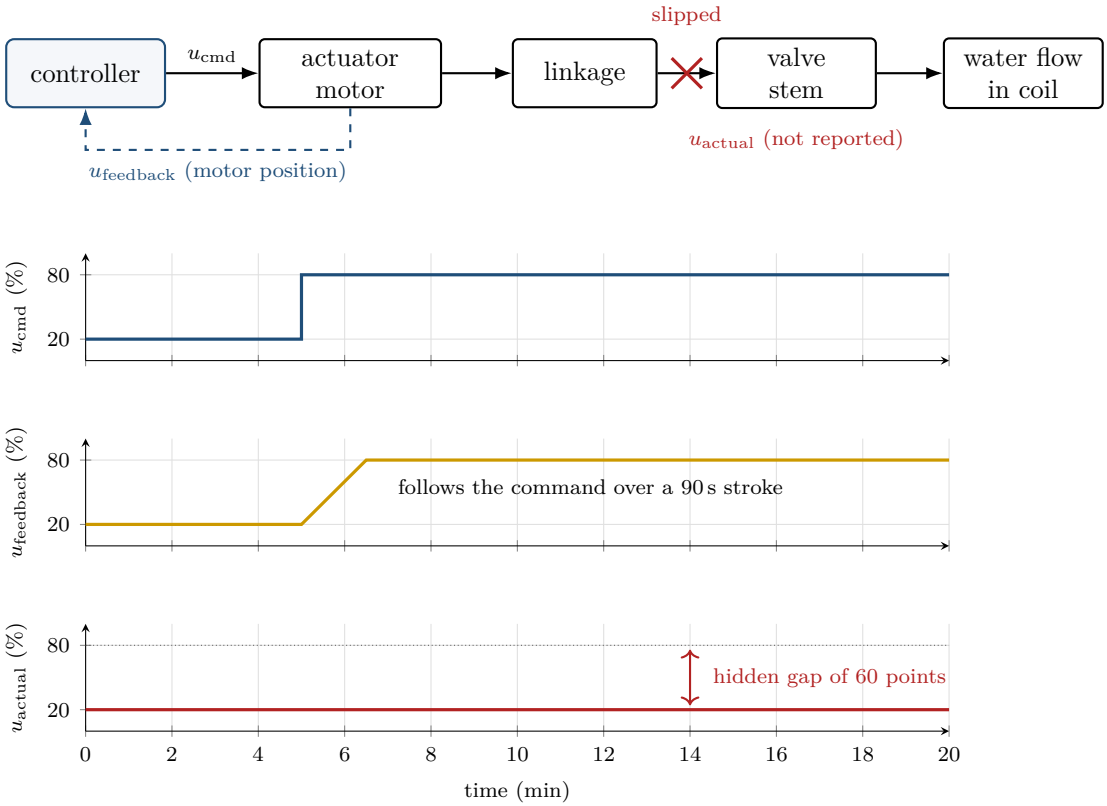


Figure 9.3: Commanded, feedback and actual valve position for a slipped linkage (illustrative). Top: the actuation chain; the feedback comes from the motor, so it follows the command. Bottom: the three signals after a command step from 20 to 80 % at minute 5. The dotted line in the lowest strip marks the commanded 80 %; the actual position stays at 20 %.

9.2.7 Server telemetry and the supervisory systems

Servers report their own telemetry through a management controller on the server board: the utilization of the processors and accelerators, u_{CPU} and u_{GPU}; the temperatures of those devices, T_{CPU} and T_{GPU}; the fan speeds N_{fan}; the input power P_{server} and, for accelerators, P_{GPU}; and memory use. This telemetry bridges computational behavior and physical behavior. Utilization explains the power through the load models of Chapter 2; fan speed is the server's own response to its inlet temperature and therefore an indirect airflow signal; processor temperature shows how close the silicon is to its limit. It is also the most numerous and the least standardized telemetry in the facility: sensor positions and calibrations differ by server model, and the server's inlet sensor sits inside the chassis, where it reads warmer than the rack's front.

The telemetry is collected by two kinds of supervisory system. A building management system covers the facility infrastructure: air handling, pumps, chillers, temperatures, valves, and their alarms; it is, broadly, the system of facility controls. Data center infrastructure management covers rack power, capacity, floor space, IT equipment, cooling and power paths, and the asset inventory; it adds visibility of the IT infrastructure and of how it is connected to the facility. In short, a BMS is close to facility controls and a DCIM is close to facility plus IT infrastructure visibility, but the boundary between the two varies by vendor and by site,

and the same quantity may appear in both with different names, units and sampling rates. Reconciling the two is often the first task of any analysis that uses them together.

Table 9.1 collects the measurement types of this section. Each row gives the quantity, where it is typically measured, and what the measurement does not give; the last column is the reason most of the rest of the chapter is needed.

Table 9.1: Measurement types in a data center: what is measured, where, and what the measurement does not give.

Quantity	Typical location	What it does not give
V, I, P, Q, S, power factor, f, E	utility service, switchgear, transformers, UPS, PDUs, busways, rack PDUs	where the power goes after the meter; the split between heat and other uses is assumed
P_{server}, u_{CPU}, N_{fan}	server management controller	airflow (fan speed is a proxy); accuracy comparable to a meter
Air temperature	server inlet and exhaust, rack inlet and exhaust, aisles, CRAH supply and return, room return, outdoor	the temperature anywhere else; the flow that set it
Water temperature	chilled- and condenser-water supply and return	heat flow, unless the water flow is also metered
RH, dew point	CRAH return or supply, room	moisture at other points; RH alone changes with temperature
Δp	underfloor to room, cold to hot aisle, filters, pumps, coils	flow, unless the resistance coefficient K is known
Water flow	chillers, pumps, some coils	flow through branches without meters
Air flow	some ducts and air handlers; rarely racks	rack airflow in most facilities
Discrete states	breakers, transfer switches, pumps, chillers, generators, valves, fans	the actual position when a mechanism has failed

9.3 Time and error

A reading is a value at a time from an imperfect instrument. This section treats the time side, sampling and clock synchronization, and the error side: accuracy, the difference between bias and noise, the limits of redundancy, calibration drift, and missing or frozen data.

9.3.1 Sampling rate

A sensor that reports every 5 minutes cannot see a transient that lasts 30 seconds. If the sensor reports an instantaneous value at each interval, the transient is caught only when a sample happens to fall inside it, with probability $30/300 = 10\,\%$, and then it appears as a single isolated value that is hard to tell from a glitch. Many meters instead report averages over 1 second or over 15 minutes, and averaging removes information in a different way: it keeps the energy and loses the shape. An illustrative 2 MW excursion lasting 30 seconds on a 10 MW load raises a 15-minute average by only

$$2\,\text{MW} \times \frac{30\,\text{s}}{900\,\text{s}} = 0.067\,\text{MW},$$

0.67 % of the load, which is within ordinary fluctuation.

The choice is a trade-off of data volume against resolution in time: one channel at one-second resolution produces 86 400 records a day, the same channel at 15-minute averages produces 96. A record is described fully only when both are stated: what was measured, and how often and

over what window. Two meters on the same circuit, one reporting instantaneous samples and the other averages, can disagree substantially during transients while both are correct.

9.3.2 Time synchronization

Telemetry from different systems is stamped by different clocks, and clocks drift. Suppose a load increase begins at 10:00:00 and the rack inlet temperature begins to rise about a minute later, as the extra heat reaches the aisle. If the clock of the power meter runs 3 minutes slow, the power increase is recorded at 10:03 while the temperature rise is recorded at 10:01. In the record, temperature rises first and power follows,

$$\text{true order: } P\uparrow \rightarrow\ T\uparrow, \qquad \text{recorded order: } T\uparrow \rightarrow\ P\uparrow,$$

which reverses cause and effect. An analysis of that record could conclude that the temperature rise preceded the load, pointing to a cooling fault, when the load was the cause.

Timestamp quality is therefore critical for dynamic diagnosis, the analysis of sequences and lags. A steady-state balance averaged over an hour tolerates a 3-minute offset; a lag estimated from a one-minute response does not. Synchronizing clocks to a common time source, recording whether a timestamp marks the start, end or middle of an averaging window, and storing times in a single time standard (coordinated universal time) are measurement requirements, not housekeeping.

9.3.3 Accuracy, precision, bias and noise

A reading of 24.2°C from a sensor specified at ±0.5°C is not an exact value. It says that the true temperature is probably within about half a kelvin of 24.2°C, and its standard form is the temperature case of equation (9.1),

$$y_T = T + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma_T^2), \tag{9.9}$$

or another error model when the error is not normal. If the ±0.5°C is read as a 95 % bound on a normal error, then $\sigma_T = 0.5/1.96 \approx 0.26\,\text{K}$; if it is a hard bound on an error spread evenly over the interval, $\sigma_T = 0.5/\sqrt{3} \approx 0.29\,\text{K}$. *Accuracy* is the closeness of readings to the true value; *precision* is the closeness of repeated readings to each other. A sensor can be precise and inaccurate: its readings cluster tightly around the wrong value.

That distinction is the distinction between noise and bias. Noise is an error with zero mean, $\mathbb{E}[\epsilon] = 0$, as in equation (9.9). Bias is a nonzero systematic offset,

$$y = T + b + \epsilon, \qquad b \neq 0. \tag{9.10}$$

The two behave differently under averaging. The mean of n readings has expectation $T + b$ and standard deviation $\sigma/\sqrt{n}$: averaging shrinks the noise and leaves the bias untouched. With $\sigma = 0.4\,\text{K}$, the mean of 12 readings has a noise standard deviation of $0.4/\sqrt{12} \approx 0.12\,\text{K}$, while a bias of $-5\,\text{K}$ is still $-5\,\text{K}$. Noise produces readings that look uncertain; bias produces readings that are stable and wrong. Bias is often the more dangerous of the two, because stable readings invite trust and because no filter can remove it.

Figure 9.4 plots the three kinds of error on one time axis. The horizontal axis is thirty days; the vertical axis is temperature. The thick black curve is the true inlet temperature of an illustrative rack, near 30°C with a slow weekly swing. The thin blue trace is a noisy but unbiased sensor; it scatters around the truth. The gray trace is a biased sensor with $b = -5\,\text{K}$; it is smooth and follows every movement of the truth, five kelvin low. The red trace is a drifting sensor whose bias starts at zero and wanders downward, as Section 9.3.6 describes. The dashed line is an alarm threshold at 27°C. The noisy sensor is above the threshold at every sample, as the truth is. The biased sensor is below it at every sample, and the drifting sensor first falls below it on day 25: both report a rack that is too hot as a rack within limits.

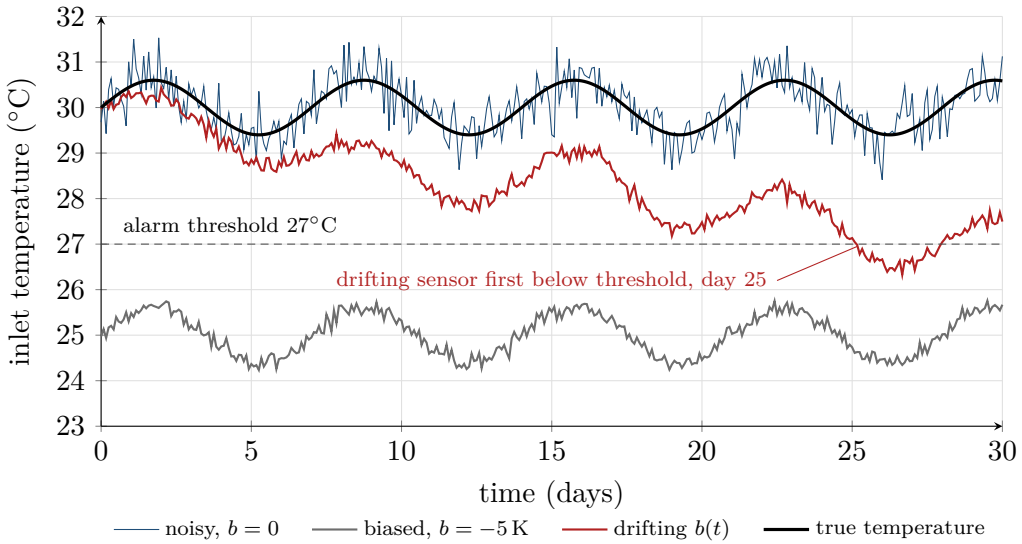


Figure 9.4: Noise, bias and drift on one time axis (illustrative, generated by `scripts/ch09_data.py`, seed 9). Truth $T = 30 + 0.6\sin(2\pi d/7)$°C sampled every 2 h; noisy sensor $y = T + \epsilon$, $\epsilon \sim \mathcal{N}(0, 0.4^2)$; biased sensor $y = T - 5 + \epsilon$, $\epsilon \sim \mathcal{N}(0, 0.1^2)$; drifting sensor $y = T + b_k + \epsilon$ with $b_{k+1} = b_k + w_k$, $w_k \sim \mathcal{N}(-0.008, 0.03^2)$ per step, reaching $b = -3.1$ K at day 30. The drift rate is exaggerated for visibility.

9.3.4 Bad sensors can look normal

Take the biased sensor of Figure 9.4 at one instant. The true inlet temperature is 30°C; the sensor has a bias of −5 K and reports 25°C. The reading is smooth, stable and within the expected range. A threshold detector set at 27°C classifies the rack as normal, and so does any check of range, rate of change or noise level: the signal has every property of a healthy sensor except that it is wrong.

Physics can say otherwise. Suppose the same rack has a metered power of $P_{\text{rack}} = 20$ kW and its fans, from their speed and the fan curve, move about 1.33 kg/s of air. Equation (9.8) predicts a temperature rise of $20/(1.33 \times 1.0) = 15$ K across the rack. If the exhaust sensor reads 45°C, the inlet implied by the other measurements is $45 - 15 = 30$°C, five kelvin above the reported 25°C. The single sensor gives no sign of trouble; the combination of four measurements, power, fan speed, exhaust and inlet, is inconsistent by 5 K. That is the value of cross-variable consistency: a fault that no check on one signal can see becomes visible when signals that obey a common physical law are compared. The comparison does not yet say that the inlet sensor is the faulty one, a point Section 9.4.3 returns to.

9.3.5 Redundant sensors help only partly

Installing several sensors at the same point is the usual defense against a faulty one. If three sensors read $T_1 = 24.1$°C, $T_2 = 24.3$°C and $T_3 = 27.0$°C, the agreement of the first two suggests that T_3 is the faulty one, and a median vote rejects it. Redundancy of this kind catches an independent failure of one sensor.

It does not catch an error the sensors share. If all three came from the same batch, were calibrated against the same faulty reference, are mounted in the same unrepresentative position, or pass through the same gateway with the same scaling error, they agree with each other and are all wrong. Write each sensor's error as a common part c shared by all and an independent

part d_i, with standard deviations σ_c and σ_d. The error of the average of n sensors has variance

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} e_i\right) = \sigma_c^2 + \frac{\sigma_d^2}{n}. \tag{9.11}$$

With $\sigma_c = \sigma_d = 0.5\,\mathrm{K}$, one sensor has an error standard deviation of $\sqrt{0.25 + 0.25} = 0.71\,\mathrm{K}$, three sensors $\sqrt{0.25 + 0.083} = 0.58\,\mathrm{K}$, and infinitely many still $0.5\,\mathrm{K}$. Adding sensors reduces only the independent part. Redundancy is not independence; a second sensor of a different type, from a different calibration chain, or a different physical relation that predicts the same quantity, is worth more than a third copy of the first.

Principle 9.2. *Averaging and redundancy remove independent errors and leave shared errors untouched. A stable, agreeing set of readings can be uniformly wrong; only an independent physical relation can expose a common bias.*

9.3.6 Calibration drift

A sensor's bias is not fixed. Thermistors age, pressure diaphragms creep, flow meters foul, and the bias changes slowly. The measurement equation becomes

$$y(t) = x(t) + b(t) + \epsilon(t), \qquad b_{t+1} = b_t + w_t, \tag{9.12}$$

in which the bias b is itself a hidden dynamic state, a random walk driven by small increments w_t. The red trace of Figure 9.4 is such a walk: its increments have a small negative mean, and over thirty days the bias wanders from zero to about $-3.1\,\mathrm{K}$, so a sensor that correctly reported an alarm at the start of the month silently stops reporting it by the end.

Realistic drift is slower, perhaps $0.1\,\mathrm{K}$ a month, or $1.2\,\mathrm{K}$ a year, but its effect depends on what the reading is used for. On a chilled-water loop whose supply-to-return difference is $5\,\mathrm{K}$, a $1.2\,\mathrm{K}$ drift in one of the two sensors changes the calculated cooling load by $1.2/5 = 24\,\%$. Because drift is a state, it can be estimated like any other state, provided there are enough independent physical constraints: a heat balance across a coil, an energy balance against metered electrical power, or a comparison with a reference during periods when two quantities are known to be equal (for example the rack inlet and the supply air when recirculation is negligible). Without such constraints a slowly drifting bias is indistinguishable from a slowly changing temperature.

9.3.7 Missing and frozen data

Telemetry has gaps. Records go missing when a network link drops packets or a gateway restarts; records are duplicated when a gateway retransmits; timestamps are invalid when a device loses its clock (a date of 1 January 1970 is a common symptom) or when a local-time clock repeats an hour at the end of daylight saving time. A missing value must also be distinguished from a zero: a power reading of $0\,\mathrm{kW}$ and the absence of a reading are different statements about a rack.

The hardest case is a frozen value. A sensor that reports exactly $24.0^{\circ}\mathrm{C}$ for six hours is either measuring a very stable temperature or no longer measuring at all, with a gateway repeating its last good value. The record alone cannot say which; context decides. A live sensor whose noise is comparable to its resolution changes its last digit within a few readings, so a perfectly flat trace while the rack's power varies by $30\,\%$ and its neighbors move by half a kelvin points to a frozen sensor. A flat trace while everything around it is also flat may be real. Data quality is a property of a reading in its physical context, not of the reading alone.

9.4 Checks, residuals and observability

Errors are found by comparing measurements with expectations. This section moves from simple plausibility rules to checks based on conservation laws, defines the residual those checks produce, and then states the structural limit that no check can overcome: a state that the measurements do not determine.

9.4.1 Plausibility checks

The simplest checks compare a reading with what the physics permits. Under normal cooling the chilled water leaves the chiller colder than it returns from the coils,

$$T_{\mathrm{CHW,s}} < T_{\mathrm{CHW,r}}, \tag{9.13}$$

for example 12°C supply and 18°C return. If the record shows the reverse, the possible explanations are swapped sensors, wrong point labels, a bad measurement, reversed flow in that part of the loop, or an unusual operating mode such as a loop being drained or a heat-recovery mode. Each calls for a different action, so the check detects a problem without naming it.

Plausibility checks of this kind are cheap and catch many errors: range checks (a rack inlet of −40°C), rate-of-change checks (a water temperature that jumps 8 K in one second), and cross-checks between discrete and continuous signals (a pump reported running with zero flow and zero pressure rise). Their limitation is the one shown in Section 9.3.4: a biased sensor that stays in range, moves slowly and agrees with the expected sign passes all of them.

9.4.2 Conservation-based checks

Stronger checks use conservation laws. Across an air-to-water heat exchanger such as a CRAH coil, the heat removed from the air must equal the heat added to the water when nothing is stored:

$$\dot{Q}_{\mathrm{air}} = \dot{m}_a \, c_{p,\mathrm{air}} \, \Delta T_a, \qquad \dot{Q}_{\mathrm{water}} = \dot{m}_w \, c_{p,w} \, \Delta T_w, \qquad \dot{Q}_{\mathrm{air}} \approx \dot{Q}_{\mathrm{water}}. \tag{9.14}$$

Take an illustrative coil with $\dot{m}_a = 40\,\mathrm{kg/s}$ of air cooled by $\Delta T_a = 20\,\mathrm{K}$, so $\dot{Q}_{\mathrm{air}} = 40 \times 1.0 \times 20 = 800\,\mathrm{kW}$, and $\dot{m}_w = 19.1\,\mathrm{kg/s}$ of water warmed by $\Delta T_w = 5\,\mathrm{K}$, so $\dot{Q}_{\mathrm{water}} = 19.1 \times 4.18 \times 5 \approx 400\,\mathrm{kW}$. The two calculations of the same heat flow differ by a factor of two.

The candidate explanations are an air flow measurement bias (if the airflow is itself estimated, a factor-two error is plausible), a water flow meter bias, a temperature bias on either side, unmodeled condensation, transient storage, and bad data mapping. Each can be examined with arithmetic. Temperature bias has the most leverage on the side with the smaller temperature difference: a bias δ in one water sensor changes $\dot{Q}_{\mathrm{water}}$ by the fraction $\delta/\Delta T_w$, so 1 K moves it by 20 %, or 80 kW, whereas 1 K on the air side moves $\dot{Q}_{\mathrm{air}}$ by 5 %. Condensation removes latent heat from the air that the sensible calculation $\dot{m}_a c_{p,\mathrm{air}} \Delta T_a$ does not count, while the water still receives it; it can therefore explain a water-side heat larger than the air-side heat, but not the reverse, and is ruled out here by its sign. Transient storage requires heat to accumulate somewhere: absorbing 400 kW for one minute would warm an illustrative 10 000 kg of water in the coil and nearby piping by $400 \times 60/(10\,000 \times 4.18) \approx 0.57\,\mathrm{K}$, which is plausible during a short transient but not for a discrepancy that persists for hours. Bad data mapping, the air readings of one CRAH paired with the water meter of another, can produce any ratio and is checked against the point list.

9.4.3 Residuals

A conservation check is expressed as a residual,

$$r = \dot{Q}_{\mathrm{air}} - \dot{Q}_{\mathrm{water}}, \tag{9.15}$$

which is close to zero when the model and the measurements are both correct. Its size is judged against its uncertainty. If the air-side calculation is uncertain by 10 % (80 kW) and the water side by 5 % (20 kW), and the two are independent, the residual has a standard deviation of $\sqrt{80^2 + 20^2} \approx 82$ kW, and the observed $r = 400$ kW is $400/82 \approx 4.9$ standard deviations from zero. The set of measurements is mutually inconsistent.

That is all one residual says. A large $|r|$ shows that the measurements cannot all be right; it does not say which one is wrong, because a bias in any of the five quantities in equation (9.14), or an error in the model, produces the same residual. Isolating the fault needs several residuals that involve different subsets of the measurements. If the IT equipment in the room served by the coil draws a metered $P_{\mathrm{IT}} = 790$ kW, then $\dot{Q}_{\mathrm{air}} - P_{\mathrm{IT}} = 10$ kW is consistent, $\dot{Q}_{\mathrm{water}} - P_{\mathrm{IT}} = -390$ kW is not, and the pattern points to the water side. The general form, measured minus predicted, applies to any measurement for which a model makes a prediction.

Principle 9.3. *A residual detects that a set of measurements is mutually inconsistent; it does not identify the faulty member. Isolation requires several residuals whose patterns differ across the candidate faults.*

9.4.4 Observability

A state is observable if the measurements contain enough information to infer it. Consider again $\dot{Q} = \dot{m} c_p \Delta T$. If $\dot{Q}$ and ΔT are both measured, for example $\dot{Q}$ from the metered electrical power of a rack and ΔT from its inlet and exhaust sensors, the mass flow follows,

$$\dot{m} = \frac{\dot{Q}}{c_p \, \Delta T}, \tag{9.16}$$

and airflow is observable. If only ΔT is measured, every combination of $\dot{Q}$ and $\dot{m}$ with the same ratio fits, and airflow is not identifiable.

Observability is a structural question, not a question of data quality. It is decided by which functions of the state are measured, the h of equation (9.1), before any data arrive. A longer record, a more accurate sensor or a better algorithm does not change it: a state that does not influence any measurement, or influences them only in a combination with another unknown, cannot be separated from that other unknown however the data are processed. The same question is asked of a transmission network in power-system state estimation, where the placement of meters decides which bus voltages can be determined before a single measurement is taken.

9.4.5 Hidden airflow

A rack with $T_{\mathrm{in}} = 20$°C and $T_{\mathrm{out}} = 35$°C has $\Delta T = 15$ K. With $c_{p,\mathrm{air}} = 1.0$ kJ/(kg K), equation (9.8) gives $\dot{Q} = 15\,\dot{m}$ (kW, with $\dot{m}$ in kg/s), and every pair on that line explains the two temperatures equally well:

$$(\dot{Q}, \dot{m}) = (15\,\mathrm{kW},\ 1\,\mathrm{kg/s}), \quad (30\,\mathrm{kW},\ 2\,\mathrm{kg/s}), \quad (60\,\mathrm{kW},\ 4\,\mathrm{kg/s}).$$

A lightly loaded rack with little air, a medium rack with moderate air, and a dense rack with a lot of air produce exactly the same inlet and exhaust temperatures.

Figure 9.5 plots that line. The horizontal axis is the mass flow through the rack; the vertical axis is the heat it carries. The solid line is the set of all pairs consistent with $\Delta T = 15$ K, and the three dots are the pairs above. The dashed horizontal line is what a metered rack power of 30 kW adds: it crosses the line at one point, $\dot{m} = 2$ kg/s, which is the observability of equation (9.16). Without that second measurement the data leave the answer anywhere on the

line. An estimation method given a prior belief about the airflow will return a point, but the point is the prior's, not the data's: along the line the measurements are equally satisfied, and no inference algorithm can recover information that is not in the measurements.

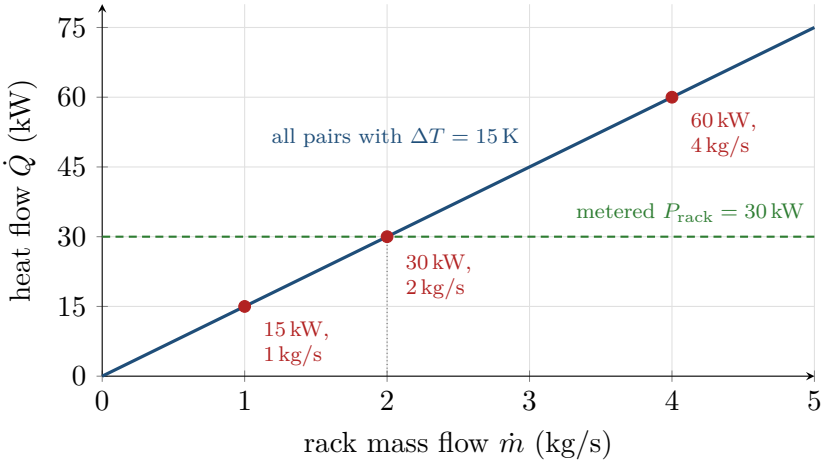


Figure 9.5: Hidden airflow: the pairs $(\dot{m}, \dot{Q})$ consistent with $T_{\text{in}} = 20°\text{C}$ and $T_{\text{out}} = 35°\text{C}$ lie on the line $\dot{Q} = \dot{m}\, c_{p,\text{air}}\, \Delta T = 15\,\dot{m}$ ($c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$, $\Delta T = 15\,\text{K}$). The dots are the three pairs of the text. A metered rack power (dashed) selects one point, $\dot{m} = 2\,\text{kg/s}$.

Principle 9.4. *No inference method recovers information that is not in the measurements. Where the data are equally satisfied along a line of states, the answer on that line comes from assumptions, and must be reported as such.*

9.5 Instrument design, placement and estimated quantities

The previous section showed what the measurements cannot do. This section turns to what can be done about it: choosing measurements for the questions they must answer, placing them where they represent what they are taken to represent, and building virtual sensors, soft sensors and fused estimates for the quantities that are not measured directly.

9.5.1 Instrumentation design follows inference goals

Instrumentation is usually installed for control and alarms, and later asked to support diagnosis. The questions it can answer are fixed at installation. Consider the operational question "is low airflow causing the hot spot at this rack?". Answering it requires measurements that bear on airflow: the rack power, which with the inlet and exhaust temperatures gives the airflow through equation (9.16); a direct airflow measurement; a pressure difference across the rack or the tile, which with a known resistance gives flow through equation (9.7); or server fan telemetry, which with the fan curves gives the server airflow. A hall with dense temperature sensing and none of these cannot answer the question, however many temperature points it has.

The design rule follows: measurement design should start from the inference goals, the questions the operator must answer and the hypotheses that must be separated, and not from a general intention to install many sensors. For each question, the designer writes down the unknowns, the relations that connect them to possible measurements, and the smallest set of measurements that makes the unknowns observable. The same budget spent on a few well-chosen flow and power points usually determines more than it would spent on additional temperature points.

9.5.2 Sensor placement and representativeness

A sensor measures its own location. In Figure 9.2, sensor 2 in the middle of the cold aisle reads 22°C, close to the supply air, while sensor 3 at the top of the facing rack reads 28°C, because hot exhaust curls over the rack and mixes into its upper inlet. Both sensors are accurate. If sensor 2 is the only cold-aisle point and is taken as the inlet temperature of the rack, the analysis is wrong by 6 K at the top of the rack, where the hottest servers sit. The error is not in the instrument but in what it is taken to represent: a representativeness problem.

With the mixing model of Chapter 6, the inlet temperature at a point is $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{hot}}$. Taking the cold-aisle reading as the supply, $T_{\text{supply}} = 22°\text{C}$, and an illustrative exhaust of $T_{\text{hot}} = 37°\text{C}$, the top-of-rack reading implies a local recirculation fraction

$$\alpha = \frac{T_{\text{in}} - T_{\text{supply}}}{T_{\text{hot}} - T_{\text{supply}}} = \frac{28 - 22}{37 - 22} = 0.4,$$

so 40 % of the air entering the top of that rack is exhaust. A single central sensor reports none of it. Representativeness is decided by placement: inlet sensors at the bottom, middle and top of each rack face, as sensor group 1 in the figure, and at the ends of each row, where recirculation around the row ends is strongest.

9.5.3 Point sensors and fields

Sensors give values at points, $T(x_i, y_i, z_i)$ at a few locations i. The room has a continuous temperature field $T(x, y, z, t)$ and a velocity field $\mathbf{v}(x, y, z, t)$ that carries heat through it, and the velocity field in particular is almost entirely unobserved. Computational fluid dynamics (CFD), the numerical solution of the equations of air flow and heat transport on a fine grid, reveals spatial structure that sparse telemetry cannot: where the recirculation zones are, which tiles feed which racks, how a blanking panel changes the flow. It does so for a modeled configuration, with assumed boundary conditions, not from measurements.

Inference from telemetry takes a different route. It does not attempt to reconstruct the whole field from a few points, which would be hopeless; it reconstructs reduced states that summarize the field for a purpose, such as one airflow and one recirculation fraction per rack. A reduced state has few enough unknowns that the available sensors can determine it, and it is the quantity the operational question needs. The price is that whatever the reduced state does not represent, a vortex behind one server or a gradient across one rack face, is invisible to the estimate.

9.5.4 Virtual sensors

A virtual sensor is an inferred quantity computed from measurements by a physical relation and published like a measurement. Two common examples are the rack airflow and the cooling load of a coil,

$$\hat{\dot{m}}_{\text{rack}} = \frac{P_{\text{rack}}}{c_{p,\text{air}}\,(T_{\text{out}} - T_{\text{in}})}, \qquad \hat{\dot{Q}} = \dot{m}_w\, c_{p,w}\, \Delta T_w. \tag{9.17}$$

In the book's terms the inputs are measured, the formula is a calculation, and the output is an inference about a state that is not measured, valid under the assumptions built into the formula.

A virtual sensor inherits uncertainty from three sources: the measurements, the assumptions and the parameters. For the measurements, take an illustrative rack with $P_{\text{rack}} = 30\,\text{kW} \pm 2\,\%$, $T_{\text{in}} = 20°\text{C}$ and $T_{\text{out}} = 35°\text{C}$, each $\pm 0.5\,\text{K}$. The temperature difference is uncertain by $\sqrt{0.5^2 + 0.5^2} = 0.71\,\text{K}$, or 4.7 % of 15 K, and the relative uncertainties of a quotient add in quadrature:

$$\hat{\dot{m}}_{\text{rack}} = \frac{30}{1.0 \times 15} = 2.0\,\text{kg/s}, \qquad \frac{\sigma_{\hat{\dot{m}}}}{\hat{\dot{m}}} = \sqrt{0.02^2 + 0.047^2} \approx 5.1\,\%,$$

so $\hat{\dot{m}}_{\text{rack}} = 2.0 \pm 0.10\,\text{kg/s}$. A bias behaves differently: if the exhaust sensor reads 1 K low, ΔT becomes 14 K and the estimate 2.14 kg/s, 7 % high, with no indication in the result. The assumptions are that all the rack's electrical power leaves as heat in the air stream (no liquid cooling, negligible conduction to the structure), that the rack is in steady state, and that one exhaust point represents the mixed exhaust, which is often stratified. The parameter is the specific heat, which varies little. For the coil, with $\dot{m}_w = 19.1\,\text{kg/s}$ and water sensors of $\pm 0.2\,\text{K}$ each on a 5 K difference, the temperature term alone contributes $\sqrt{0.2^2 + 0.2^2}/5 \approx 5.7\,\%$. A virtual sensor is as credible as its weakest input and its weakest assumption, and should be published with its uncertainty.

9.5.5 Soft sensors

A soft sensor goes further: it combines measurements, empirical models, machine learning and physical constraints into an estimate. A rack heat soft sensor might take the form

$$\hat{\dot{Q}}_{\text{rack}} = f(P_{\text{PDU}},\ T_{\text{in}},\ T_{\text{out}},\ N_{\text{fan}}), \tag{9.18}$$

where f is fitted to periods with a trusted reference, for example a campaign with a portable airflow hood, and is constrained by physics: the heat is nonnegative, it approaches the electrical power in steady state, and the airflow implied by N_{fan} scales with fan speed as the fan laws require. The physical constraints keep the estimate plausible where the data are thin.

A soft sensor is an estimate, not a measurement. Its error has a component from the model form and the training data, and it can be large where the operating point lies outside the range the model was fitted on, such as a new server type, a changed containment or a fan failure. Its output is labeled as estimated in the telemetry it feeds, and it is revalidated when the equipment it describes changes.

9.5.6 Sensor fusion

The airflow through a rack bears on several measurements at once: the power and temperatures through the energy balance, a pressure difference through the flow resistance, and the fan speed through the fan curve. Sensor fusion uses all of them together. In probabilistic form, the fused estimate is the distribution of the airflow given all the readings,

$$p(\dot{m} \mid P_{\text{rack}}, T_{\text{in}}, T_{\text{out}}, \Delta p, N_{\text{fan}}) \ \propto\ p(\dot{m})\, p(P_{\text{rack}}, T_{\text{in}}, T_{\text{out}} \mid \dot{m})\, p(\Delta p \mid \dot{m})\, p(N_{\text{fan}} \mid \dot{m}), \tag{9.19}$$

where the product of likelihoods assumes that the three groups of measurements err independently given the airflow.

When each group gives an approximately normal estimate, fusion reduces to an inverse-variance weighted mean. With illustrative estimates $2.0 \pm 0.10\,\text{kg/s}$ from the energy balance, $1.8 \pm 0.20\,\text{kg/s}$ from fan speed and $1.9 \pm 0.15\,\text{kg/s}$ from the pressure difference, the weights are $1/0.10^2 = 100$, $1/0.20^2 = 25$ and $1/0.15^2 = 44.4$, and

$$\hat{\dot{m}} = \frac{100 \times 2.0 + 25 \times 1.8 + 44.4 \times 1.9}{100 + 25 + 44.4} = \frac{329.4}{169.4} \approx 1.94\,\text{kg/s}, \qquad \sigma = \frac{1}{\sqrt{169.4}} \approx 0.077\,\text{kg/s}.$$

The fused estimate is more precise than the best single one, and it is more robust: the three estimates rest on different sensors and different physical relations, so a bias in one shows as a disagreement among them instead of passing silently into the result. That robustness depends on the independence assumption of equation (9.19); estimates that share a sensor share its bias, as in equation (9.11).

9.6 From raw sensors to risk

This section organizes the chapter's objects into a hierarchy from raw readings to a probability of a limit being exceeded, separates the capabilities each layer provides, shows how to choose the next measurement and what it costs, and classifies the variables of an operational question by what the instrumentation can say about them.

9.6.1 The instrumentation hierarchy

The quantities of the chapter fall into five layers. Layer 1 is the raw sensors: V, I, T, p, $\dot{m}$, RH. Layer 2 is derived quantities, calculated from raw readings by known relations, such as electrical power $P = VI\cos\varphi$ and heat flow $\dot{Q} = \dot{m}c_p\Delta T$. Layer 3 is estimated physical states and parameters that are not measured: the rack airflow $\dot{m}_{\text{rack}}$, the coil conductance UA, the state of health SOH of a UPS battery. Layer 4 is diagnosis, a probability of each hypothesis given the measurements, $\Pr(H_i \mid y)$. Layer 5 is risk, the probability that a limit will be crossed, such as $\Pr(T > T_{\max})$ or $\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}})$.

Figure 9.6 draws the hierarchy. The five boxes are the layers, from raw sensors at the bottom to risk at the top, and each arrow is the step from one layer to the next. The brace on the left groups the layers by the book's three kinds of statement: layer 1 is measured, layer 2 calculated, and layers 3 to 5 inferred. The notes on the right say what each step needs in addition to the layer below. The higher a quantity sits, the more it depends on models and assumptions, and the more its credibility rests on everything beneath it.

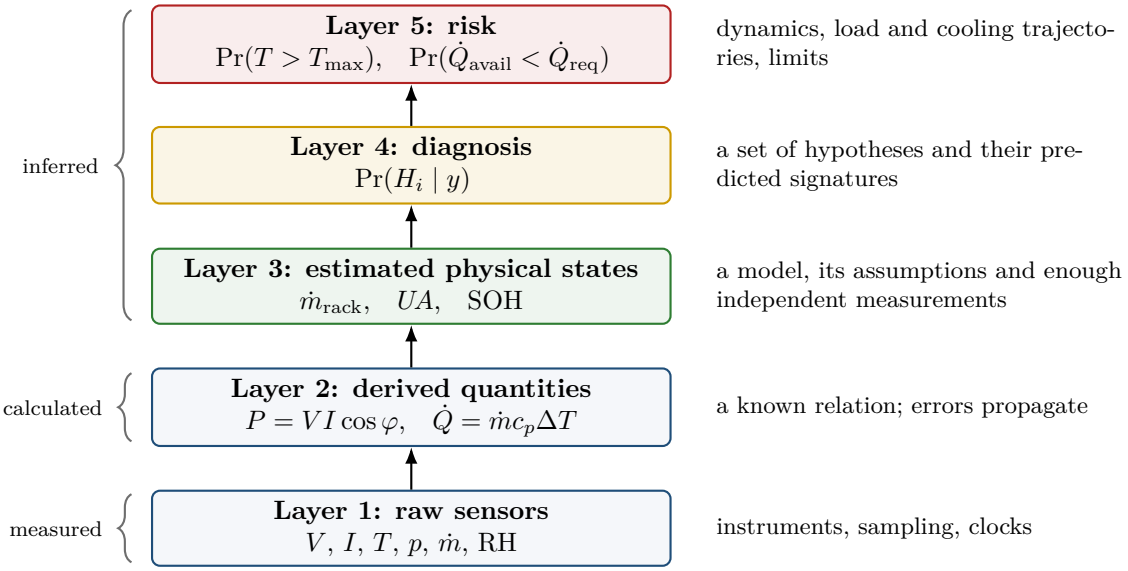


Figure 9.6: The instrumentation hierarchy from raw sensors to risk. Each layer is built on the one below and needs what the note on its right names; layer 1 is measured, layer 2 calculated, and layers 3 to 5 inferred.

9.6.2 Detection, diagnosis and prediction

The layers correspond to distinct capabilities, and a facility can have one without the next. A single rack inlet temperature sensor can detect that something is wrong: the reading rises above its usual range. It cannot say why. A high IT load, low airflow from the floor, recirculation of hot exhaust, and a bad sensor all raise the reading. Each cause leaves a different signature in other variables, a higher rack power for the load, a lower tile pressure for the airflow, a

top-to-bottom gradient for recirculation, no change in any neighbor for a bad sensor, but the one sensor sees none of them. Detection capability is not diagnostic capability.

Diagnosis is not prediction either. Knowing that the airflow is low does not give the time until the rack reaches its thermal limit. That needs the thermal capacitance of the air and equipment (Chapter 8), the trajectory of the load, the trajectory of the airflow, and the behavior of the cooling system as it responds. With an illustrative effective capacitance of 600 kJ/K for an aisle and its equipment, an inlet at 27°C and a limit of $T_{\max} = 32$°C, a net heat imbalance of 10 kW warms the aisle at $10/600\,\mathrm{K/s} = 1\,\mathrm{K/min}$ and reaches the limit in about 5 minutes; an imbalance of 2 kW takes about 25 minutes. The same diagnosis supports two very different operating responses. Whether an instrumentation set is adequate therefore depends on the downstream question: detection, diagnosis and prediction each need more than the one before.

Principle 9.5. *Detecting an anomaly, diagnosing its cause and predicting its consequence are three capabilities with increasing measurement requirements; an instrumentation set adequate for one may be inadequate for the next.*

9.6.3 The next-best measurement

When two hypotheses explain the current data equally well, the useful next measurement is the one that would tell them apart. Suppose a rack inlet sensor reads 30°C while its neighbors read 24°C, and two hypotheses remain: H_1, low airflow at the rack, so the inlet really is 30°C; and H_2, sensor bias, so the inlet is 24°C and the sensor reads 6 K high. Candidate measurements are the rack power, a server exhaust temperature, the server fan speed, a direct airflow measurement at the tile, and a second, independent inlet sensor. The best of them is the one whose predicted distribution under H_1, $p(y_{\text{new}} \mid H_1)$, differs most from its predicted distribution under H_2, $p(y_{\text{new}} \mid H_2)$. For normal predictions with means μ_1 and μ_2 and a common standard deviation σ, a simple measure of the difference is the separation

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}. \tag{9.20}$$

The expected reduction in uncertainty about the hypotheses that a measurement provides is its value of information; for two hypotheses it grows with d.

Table 9.2 lists illustrative predictions. Each row is one candidate; the second and third columns are its expected reading under each hypothesis, the fourth the measurement's standard deviation, and the last the separation d. The rack power barely moves under either hypothesis (only the extra fan power under H_1), so it carries almost no information about this question, $d \approx 0.75$. The second inlet sensor separates them by twelve standard deviations, and the fan speed and server exhaust also discriminate well, because under H_1 the servers really are breathing hot air and respond to it. A second inlet reading of 29.6°C would make H_1 more likely than H_2 by a likelihood ratio of about e^{62}; the question would be settled by one inexpensive sensor.

Table 9.2: Candidate next measurements for separating low airflow (H_1) from inlet sensor bias (H_2); illustrative predictions.

Candidate	under H_1	under H_2	σ	d
Second inlet sensor	30°C	24°C	0.5 K	12
Server fan speed	8000 rev/min	6000 rev/min	300 rev/min	6.7
Server exhaust temperature	42°C	36°C	1.0 K	6
Direct airflow at the tile	$0.6\,\mathrm{m^3/s}$	$1.0\,\mathrm{m^3/s}$	$0.1\,\mathrm{m^3/s}$	4
Rack power	20.3 kW	20.0 kW	0.4 kW	0.75

9.6.4 Instrumentation economics

Every measurement costs more than its purchase price. It must be installed, often in an operating hall where access is restricted; calibrated at intervals; connected to a network and integrated into the BMS or DCIM, with a point name that someone must keep correct; maintained and replaced when it fails; its data stored, at a volume set by its sampling rate; secured, because every networked device is a potential entry point; and its failures managed, because a failed sensor produces alarms, gaps and, as Section 9.3.4 showed, sometimes plausible wrong values. Over a sensor's life these costs usually exceed its purchase price several times over.

The design objective is therefore not to measure everything. It is to measure enough to answer the decisions that matter, choosing each measurement for the information it adds to a named question, as in Table 9.2, and weighing that information against its cost. A second inlet sensor that settles the airflow-versus-bias question for a few hundred dollars of installed cost is worth more than a hundred further temperature points that add nothing to it.

9.6.5 A classification of variables

For one operational question, every variable involved falls into one of four classes:

- **directly observed**, y: a measured quantity, such as P_{rack}, T_{in} and T_{out};
- **latent but inferable**, x_{latent}: a state that is not measured but is determined by the measurements under stated assumptions, such as the rack airflow from the energy balance;
- **a parameter requiring estimation**, θ: a constant or slowly varying quantity that the data can determine over time, such as the normal airflow of the rack or the bias of a sensor;
- **fundamentally unidentifiable with the current instrumentation**, $z_{\text{unidentifiable}}$: a quantity whose possible values produce the same measurements, such as which of two mechanisms caused an airflow drop.

The classification belongs to the question and the instrumentation together: adding a sensor can move a variable from the fourth class to the second.

Figure 9.7 draws the classification for the hot-spot question. The box at the top states the question and the instrumentation. The measurements on the left feed an inference step, which produces the distribution of the unknowns given the data. The three boxes on the right are what the inference returns for each unknown class: for the latent airflow and the parameters the distribution narrows, and the result is a statement that the dataset supports this inference; for the unidentifiable variable the distribution after the data is the same as before, and the result is the statement that these two hypotheses cannot be distinguished with the current measurements. Both statements are results of the analysis, not failures of it: an analysis that can say "the data do not separate these" is more useful than one that always returns an answer.

9.6.6 An example capability map

A capability map states, for a given instrumentation set, which questions it can answer and under what assumptions. Take a rack with three measurements: P_{rack}, T_{in} and T_{out}. The rack airflow $\dot{m}_{\text{rack}}$ can be estimated by equation (9.17), under the assumptions that the electrical power leaves as heat in the air and that the two temperature points represent the inlet and the mixed exhaust. Airflow degradation can be detected: a sustained fall of the estimate at constant power, against the rack's own normal airflow, is a degradation whatever its cause. A blocked filter cannot be separated from fan degradation: both reduce the airflow, and with these three measurements both produce the same fall of $\hat{\dot{m}}_{\text{rack}}$ and the same rise of ΔT.

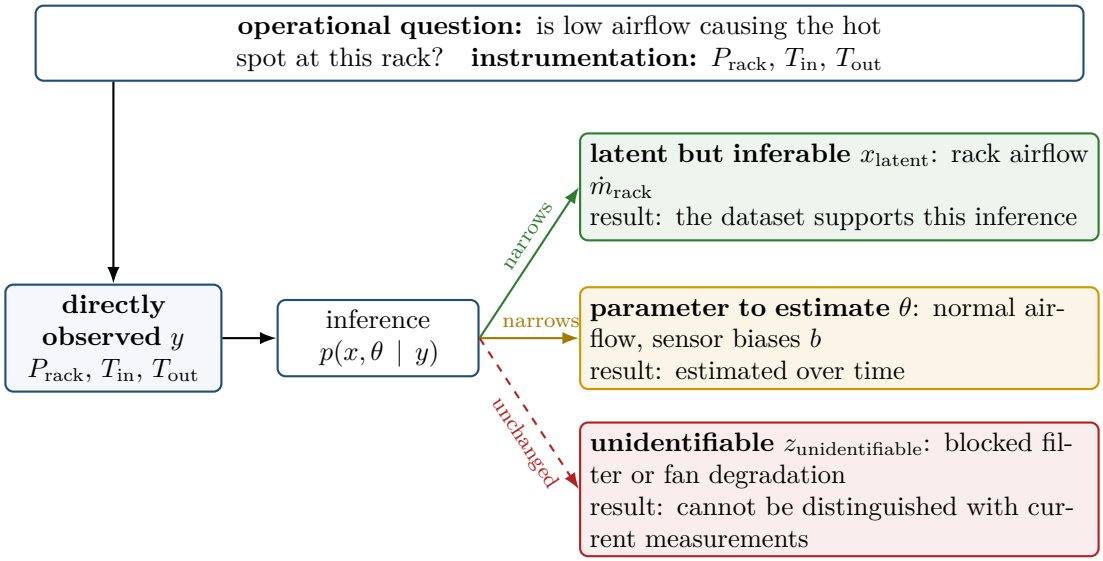


Figure 9.7: The four-way classification of variables for one operational question. Directly observed variables enter the inference; latent states and parameters that the data determine come out with narrower distributions; a variable the instrumentation cannot resolve comes out as it went in, and the analysis reports that as its result.

Separating them needs a measurement on which the two mechanisms act differently. A pressure difference across the filter rises when the filter is blocked and falls when the fans weaken; a fan speed falls when a fan degrades but rises when the fans speed up against a blocked filter. Either Δp or N_{fan} therefore moves the third question from unidentifiable to inferable.

Figure 9.8 draws the map. The three measurement boxes on the left feed the energy-balance virtual sensor in the middle, which supports the two capabilities in green on the right. The red box is the capability the three measurements cannot provide. The two dashed boxes at the lower left are measurements not installed; the dashed arrows show that either would provide it.

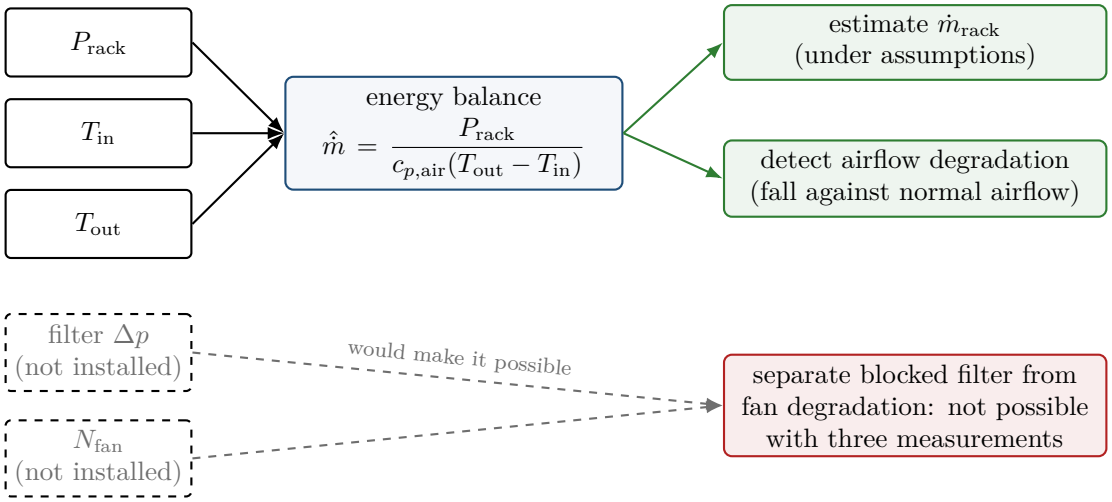


Figure 9.8: Capability map for a rack instrumented with P_{rack}, T_{in} and T_{out}. The energy-balance virtual sensor supports airflow estimation and degradation detection; separating a blocked filter from fan degradation needs a filter pressure difference or fan speed (dashed).

9.6.7 Measured, inferable and not inferable

Four equations carry the chapter:

$$\text{measurement:} \quad y = h(x, \theta) + \epsilon, \tag{9.21}$$

$$\text{bias:} \quad y = x + b + \epsilon, \tag{9.22}$$

$$\text{heat and flow:} \quad \dot{Q} = \dot{m}\, c_p\, \Delta T, \tag{9.23}$$

$$\text{residual:} \quad r = y_{\text{measured}} - y_{\text{predicted}}. \tag{9.24}$$

The first says what a sensor reports; the second, how it is most dangerously wrong; the third, why temperature without flow does not give heat; the fourth, how inconsistencies are found.

The deeper statement is that a data center does not observe its state. It observes partial, noisy and sometimes biased evidence of it,

$$\text{true physical system} \rightarrow \text{sensors} \rightarrow \text{incomplete evidence} \rightarrow \text{inference},$$

and every quantity of interest is in one of three situations: measured, inferable from the current data under stated assumptions, or not inferable from the current data. The inlet temperature at a sensor is measured; the rack airflow, with metered power, is inferable; the cause of an airflow drop, without a pressure or fan-speed signal, is not. Keeping the three apart is what prevents a model output from being reported as a claim the measurements cannot support.

9.7 What to retain

Definition 9.1 (Measurement architecture)**.** A sensor reports $y = h(x, \theta) + \epsilon$, a function of the state with noise, bias and drift, at sampled and imperfectly stamped times. What a data center can know is bounded by which functions are measured: some states are measured, some are inferable from the measurements under stated assumptions, and some are not inferable from the current data, and an analysis reports which is which.

Table 9.3 lists the chapter's objects and the subsection that defines each.

9.8 Exercises

Exercise 9.1. A facility's utility meter reads 8.4 MW and its PDU meters sum to 6.9 MW. Sub-meters read 1.05 MW for cooling and pumps, 0.28 MW for transformer and UPS losses, and 0.06 MW for lighting and auxiliaries. (a) Compute the unexplained difference without sub-meters and the residual of equation (9.6) with them, in kilowatts and as a percentage of P_{facility}. (b) If each meter is accurate to 0.5 %, is the residual consistent with correct meters? (c) A week later the residual is 210 kW while all sub-meters are unchanged. List three explanations and the check that would test each.

Exercise 9.2. An inlet sensor has noise $\sigma = 0.3$ K and bias $b = +1.2$ K, and the true inlet temperature is 26.5°C against an alarm threshold of 27°C. (a) What is the mean reading, and what is the standard deviation of the average of 16 readings? Does the averaged reading raise an alarm, and is the alarm correct? (b) Three such sensors share a common error with $\sigma_c = 0.4$ K and have independent errors with $\sigma_d = 0.3$ K. Compute the error standard deviation of one sensor, of the average of three, and of the average of infinitely many, using equation (9.11).

Exercise 9.3. A CRAH coil has $\dot{m}_a = 30$ kg/s of air cooled by 12 K and $\dot{m}_w = 12$ kg/s of water warmed by 6.5 K. (a) Compute $\dot{Q}_{\text{air}}$, $\dot{Q}_{\text{water}}$ and the residual r. (b) With uncertainties of 8 % on the air side and 4 % on the water side, how many standard deviations is r from zero? (c) A 0.5 K bias appears in the chilled-water return sensor, reading high. Recompute $\dot{Q}_{\text{water}}$ and

Table 9.3: The objects of the measurement architecture and where each is defined.

Object	Expression	Role	Defined in
Measurement equation	$y = h(x, \theta) + \epsilon$	what a sensor reports	9.1.2
Metering hierarchy	P_{facility}, P_{UPS}, P_{PDU}, P_{rack}, P_{server}	parent meters check children	9.2.1
Facility balance	equation (9.6)	a consistency check on power	9.2.2
Pressure and flow	$\Delta p = K\dot{V}^2$	flow only through an uncertain K	9.2.4
Heat and flow	$\dot{Q} = \dot{m} c_p \Delta T$	ΔT is half of a heat measurement	9.2.5
Commanded, feedback, actual	u_{cmd}, u_{feedback}, u_{actual}	hidden actuator state	9.2.6
Noise and bias	$y = T + b + \epsilon$	averaging removes noise, not bias	9.3.3
Common-mode error	equation (9.11)	redundancy is not independence	9.3.5
Drift	$b_{t+1} = b_t + w_t$	bias as a hidden state	9.3.6
Residual	$r = \dot{Q}_{\text{air}} - \dot{Q}_{\text{water}}$	detects inconsistency, not its source	9.4.3
Observability	$\dot{m} = \dot{Q}/(c_p \Delta T)$	a structural property	9.4.4
Hidden airflow	line $\dot{Q} = 15\,\dot{m}$	states the data cannot separate	9.4.5
Virtual sensor	equation (9.17)	a calculated estimate of a state	9.5.4
Soft sensor	equation (9.18)	a model-based estimate	9.5.5
Sensor fusion	equation (9.19)	several relations, one estimate	9.5.6
Five layers	Figure 9.6	raw sensors to risk	9.6.1
Separation	$d = \lvert\mu_1 - \mu_2\rvert/\sigma$	choosing the next measurement	9.6.3
Four classes	y, x_{latent}, θ, $z_{\text{unidentifiable}}$	what the data say about each variable	9.6.5

r. (d) The IT equipment served by the coil draws 330 kW. Using the three pairwise residuals among $\dot{Q}_{\text{air}}$, $\dot{Q}_{\text{water}}$ and P_{IT} after the bias, say which side is suspect.

Exercise 9.4. A rack has $T_{\text{in}} = 21°\text{C}$ and $T_{\text{out}} = 33°\text{C}$. (a) List the heat flows corresponding to airflows of 0.5, 1.5 and 3 kg/s, and state why the temperatures alone cannot select among them. (b) The rack power is metered at $18\,\text{kW} \pm 2\,\%$ and each temperature is uncertain by $\pm 0.4\,\text{K}$. Compute $\hat{\dot{m}}_{\text{rack}}$ and its standard deviation with equation (9.17). (c) A fan-curve estimate gives $1.4 \pm 0.15\,\text{kg/s}$. Fuse the two estimates by inverse-variance weighting and give the fused mean and standard deviation. (d) Classify T_{in}, $\dot{m}_{\text{rack}}$, the rack's normal airflow and the cause of a future airflow drop into the four classes of Section 9.6.5.

9.9 Connection to the TERRA framework

The chapter's objects are the measurement side of a graph model: they say how the conservation networks of the earlier chapters are attached to data, and what the data can and cannot determine about them.

- **Each sensor is a measurement attached to a node or an edge.** A rack inlet temperature is a measurement of a state at a rack node, a PDU meter a measurement of the power on an electrical edge, a water flow meter a measurement of the mass flow on a loop edge. Equation (9.1) is the likelihood that connects each reading to the graph, with its noise model, and the measurement types of Table 9.1 are the kinds of attachment available.

- **Balances are residual generators.** The facility balance of equation (9.6) is conservation at the root of the electrical network, and the air-against-water balance of equation (9.14) is conservation of energy at a coil node coupling the air and water networks. Each closes or fails to close as a residual, and the pattern across several balances, as in Section 9.4.3, isolates the faulty measurement.

- **Bias, drift and actuator position are states.** The bias of equation (9.12) enters the state vector as a slow random walk, and the actual valve position of Figure 9.3 as a component state distinct from its command. Both are inferred from the balances they disturb, and status points that set breakers and valves make the topology itself an uncertain object.

- **Inferred quantities are posteriors, and unidentifiable ones are reported.** Rack airflow, recirculation fraction and coil conductance are inferred, with posteriors that narrow when the measurements constrain them. Along the line of Figure 9.5 the posterior equals the prior, and the four-way classification of Section 9.6.5 is the framework's report for a query: observed, inferable, estimated, or not distinguishable with the current instrumentation.

- **Diagnosis, risk and the next measurement come from the same posterior.** Layer 4 of Figure 9.6 is a ranked posterior $\Pr(H_i \mid y)$ over hypotheses for one symptom; layer 5 is the probability that available cooling is below required cooling, $\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}})$, from forward simulation; and the next measurement is the one whose predicted distributions differ most across the hypotheses, as in Table 9.2.

In summary, a data center observes partial, noisy and sometimes biased evidence of its state through sensors attached at a few nodes and edges of its conservation networks; balances turn that evidence into residuals, inference turns residuals and readings into posteriors over hidden states and hypotheses, and the structure of the measurements decides, before any data arrive,

which of those posteriors can narrow at all. The next chapter, Chapter 10, turns from observing the system to acting on it: the control loops that read these sensors and drive the actuators whose commanded and actual positions this chapter separated.

Chapter 10

Controls: feedback loops, staging, predictive control and action under uncertainty

10.1 The control problem

The previous chapters described what a data center is made of and what can be known about it: the power path, the heat the IT equipment produces, the air and water that carry that heat away, the plant that rejects it, the time scales on which temperatures respond (Chapter 8) and the sensors that report all of it (Chapter 9). None of those chapters changed anything in the facility. This chapter connects the measured quantities to the quantities the facility can manipulate. Its question is

given the measured temperatures, pressures, flows, loads and equipment states, how should a data center act so that it stays safe and runs efficiently?

The answer has the form of a loop: measure, compare the measurement with what is wanted, act, and measure again. A data center runs that loop at many levels at the same time, from a fan controller that corrects its speed every second to a supervisory layer that decides every few minutes which chillers should run and how much risk the facility is carrying. The chapter builds the loop from the simplest feedback controller up to control under uncertainty, in which the controller acts on a probability distribution over the state rather than on a single measured value.

Jargon: setpoint, actuator, negative feedback, hysteresis, PID, cascade, staging, model predictive control, chance constraint, robust control

A *setpoint* is the value a controller tries to hold a measured variable at, for example 24°C for a rack inlet temperature. An *actuator* is the device through which a controller changes the physical system: a fan motor, a valve, a pump motor, a compressor. *Negative feedback* is the arrangement in which the controller's action opposes the deviation it measures, so that a rise in temperature produces more cooling and a fall produces less. *Hysteresis* in a controller is a deliberate gap between the threshold that switches an action on and the threshold that switches it off, so that a measurement hovering near one value does not make the actuator switch back and forth. *PID* stands for proportional, integral and derivative: a controller whose action is the sum of a term proportional to the present error, a term proportional to its accumulated integral and a term proportional to its rate of change. A *cascade* is a pair of nested loops in which an outer controller computes the setpoint of an inner controller. *Staging* is the switching of whole units on and off, here the number of chillers running, as a discrete decision. *Model predictive control* (MPC) predicts the system over a future horizon with a model, chooses the best sequence of actions over that horizon, applies only the first and repeats at the next step. A *chance constraint* requires a condition to hold with at least a stated probability rather than with certainty. *Robust control* chooses actions that keep the system acceptable for every value of an uncertain parameter within a declared range.

Key idea

Control is the choice of the inputs u that make the facility behave as wanted. Simple feedback loops do this well when the state is measured and the disturbances are small; when the state, the parameters, the loads and the future disturbances are uncertain, the choice of u requires state estimation, forecasting, optimization and an explicit statement of risk.

This section writes the facility as a dynamic system with a state, a control input and a disturbance, and lists the actuators through which the control input reaches the physical equipment.

10.1.1 The dynamic system

Every control question in this chapter refers to one model,

$$\dot{x} = f(x, u, d, \theta), \tag{10.1}$$

where x is the state, the set of quantities whose present values, together with the inputs, determine the future of the system; u is the control input, the vector of quantities the facility chooses; d is the disturbance, the vector of quantities that act on the system but are not chosen; and θ is the vector of parameters, quantities such as thermal resistances, coil conductances and equipment capacities that change slowly or not at all but are not known exactly. In this chapter u always denotes the control input, not a processor utilization.

For the cooling system an illustrative choice of the three vectors is

$$x = \left[\, T_{\text{rack}},\ T_{\text{CHW,s}},\ T_{\text{CHW,r}},\ p_{\text{plenum}},\ \dot{m}_w \,\right]^{\mathsf{T}}, \tag{10.2}$$

$$u = \left[\, N_{\text{fan}},\ N_{\text{pump}},\ u_{\text{valve}},\ T_{\text{CHW,set}},\ N_{\text{ch}} \,\right]^{\mathsf{T}}, \tag{10.3}$$

$$d = \left[\, P_{\text{IT}},\ T_{\text{amb}},\ \text{RH},\ \text{equipment failures} \,\right]^{\mathsf{T}}. \tag{10.4}$$

The state lists a rack temperature T_{rack}, the chilled-water supply and return temperatures $T_{\text{CHW,s}}$ and $T_{\text{CHW,r}}$, the underfloor or supply plenum pressure p_{plenum} and the chilled-water mass flow $\dot{m}_w$. The control input lists the fan speed N_{fan} of the room air handlers, the pump speed

N_{pump}, the opening of a coil valve u_{valve} between 0 (closed) and 1 (fully open), the chilled-water supply setpoint $T_{\text{CHW,set}}$ and the number of chillers running N_{ch}. The disturbance lists the IT power P_{IT}, which becomes heat, the outdoor temperature T_{amb} and relative humidity RH, which set how easily the plant rejects heat, and equipment failures, which remove capacity without warning.

The three kinds of statement that this book keeps apart apply to the state. Some components are measured directly: $T_{\text{CHW,s}}$ and $T_{\text{CHW,r}}$ come from immersion sensors, p_{plenum} from a pressure transducer. Some are calculated from measurements through a known relation, such as the heat removed by a coil from its water flow and temperature difference. Others are inferred, because no sensor reports them: the airflow through a particular rack, the fraction of hot exhaust that recirculates to an inlet, the bias of a drifting sensor. The control problem is to choose $u(t)$ so that $x(t)$ stays inside its limits while the cost of operation is as low as possible, knowing x only through measurements y and knowing d and θ only partly.

10.1.2 The actuators

The control input reaches the physical system through six classes of actuator, listed in Table 10.1. Each row gives the symbol used in this chapter, the device, the physical quantity the actuator moves directly, and the variable whose control is its usual purpose.

Table 10.1: The main actuators of a data center cooling system.

Input	Actuator	Moves directly	Usually controls
$u_1 = N_{\text{fan}}$	CRAH or CRAC fans	room airflow $\dot{m}_{\text{air}}$	rack inlet temperature, plenum pressure
$u_2 = u_{\text{valve}}$	cooling-coil valve	chilled-water flow through one coil	supply air temperature of that unit
$u_3 = N_{\text{pump}}$	chilled-water pumps	loop flow $\dot{m}_w$ and differential pressure Δp	pressure available to the coil valves
u_4	chiller compressor loading	refrigerant flow, $\dot{Q}_{\text{evap}}$	chilled-water supply temperature $T_{\text{CHW,s}}$
$u_5 = N_{\text{ch}}$	chiller staging	number of chillers running	capacity and margin of the plant
u_6	cooling-tower fans	airflow through the tower	condenser-water temperature

The list makes one point plain: there is no single “cooling control”. There are many loops, and they interact because they act on shared physical quantities. The air handler fans and the coil valves both set the temperature of the supply air, one through the airflow and the other through the water flow in the coil. The coil valves and the pumps both set the water flow, one by changing the resistance of a branch and the other by changing the pressure across all branches. The chillers set the chilled-water temperature that every coil sees, and the tower fans set the condenser temperature that every chiller sees. A change made by one loop is a disturbance to the others, and much of the chapter is about what happens when loops that are each correct on their own are combined.

10.2 Feedback controllers

This section builds the single feedback loop: the comparison of a measurement with a setpoint, the sign conventions of the error, and the four controller laws used in data center cooling (on/off,

proportional, proportional–integral and proportional–integral–derivative), ending with why a slow thermal plant makes their tuning matter.

10.2.1 The simplest temperature loop

Consider a rack inlet temperature with setpoint $T_{\text{set}} = 24°\text{C}$ and a measured value of $T = 27°\text{C}$. The controller computes an error, the difference between the measurement and the setpoint, and two sign conventions for it are in use. The general control literature defines the error as reference minus measurement,

$$e = T_{\text{set}} - T = 24 - 27 = -3\,\text{K}, \tag{10.5}$$

so that a positive error means the variable is below its setpoint. That convention suits a heating loop, where a positive error calls for more of an action that raises T. A cooling actuator lowers T, so with this convention a cooling controller must be *reverse acting*: its output must rise when the error becomes more negative. Cooling practice therefore commonly writes the error the other way round,

$$e = T - T_{\text{set}} = 27 - 24 = +3\,\text{K}, \tag{10.6}$$

so that a positive error means "too warm" and calls for more cooling, and a controller with a positive gain acts in the right direction. The two conventions describe the same loop and differ only in the sign of the gain. **This chapter uses $e = T - T_{\text{set}}$, equation (10.6), throughout.**

The action is to raise the fan speed. The physical chain is

$$N_{\text{fan}} \uparrow \implies \dot{m}_{\text{air}} \uparrow \implies \dot{Q}_{\text{removed}} \uparrow \implies T \downarrow,$$

because the heat an air handler removes is $\dot{Q}_{\text{removed}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, (T_{\text{return}} - T_{\text{supply}})$. With illustrative values $\dot{m}_{\text{air}} = 4\,\text{kg/s}$, $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$ and a 12 K difference between return and supply air, the unit removes $4 \times 1.0 \times 12 = 48\,\text{kW}$; at 5 kg/s and the same temperature difference it removes 60 kW. If the heat produced in the zone has not changed, the extra 12 kW of removal makes the zone temperature fall.

Figure 10.1 draws the loop as a block diagram. Signals flow along the arrows from left to right in the forward path and from right to left in the feedback path below it. The circle is the comparison: the measured temperature enters with a plus sign and the setpoint with a minus sign, so its output is the error of equation (10.6). The controller turns the error into a command u; the actuator, here the air handler fan, turns the command into airflow and cooling; the plant, the room and the rack, turns the balance of heat into a temperature. The IT power enters the plant from above as a disturbance. The sensor closes the loop by reporting T back to the comparison.

Why negative feedback works. The loop is stable in direction because the action opposes the deviation that caused it. If the temperature rises, the error grows, the cooling effort grows and the temperature is pushed down; if the temperature falls, the error shrinks or turns negative, the cooling effort falls and the temperature is allowed to rise. A positive-feedback arrangement, in which a rise in temperature reduced the cooling, would amplify every deviation. The sign of the error in equation (10.6) and the sign of the gain are what make the feedback negative, which is why the choice of convention must be stated and kept.

> **Principle 10.1.** *A feedback loop measures, compares with the setpoint and acts against the deviation; with $e = T - T_{\text{set}}$, more cooling for a positive error is negative feedback, and the loop resists every disturbance it can see.*

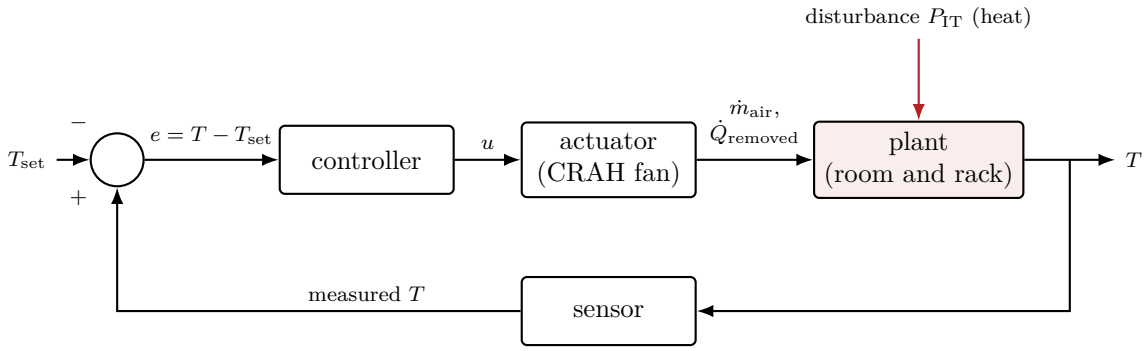


Figure 10.1: The simplest temperature loop as a block diagram. The comparison forms the error $e = T - T_{\mathrm{set}}$ (the convention of this chapter); the controller converts it into a command u; the fan converts the command into airflow and heat removal; the room and rack convert the heat balance, including the IT power entering as a disturbance, into the temperature T, which the sensor reports back.

10.2.2 On/off control with hysteresis

The simplest controller law has two values. The cooling is either on or off:

$$u(t) = \begin{cases} 1 & \text{if } T > T_{\mathrm{high}}, \\ 0 & \text{if } T < T_{\mathrm{low}}, \\ u(t^-) & \text{otherwise}, \end{cases} \tag{10.7}$$

where $u(t^-)$ means the previous value is kept. For example, the fan turns on above $T_{\mathrm{high}} = 26°\mathrm{C}$ and off below $T_{\mathrm{low}} = 24°\mathrm{C}$. The 2 K gap between the two thresholds is the hysteresis band. Between 24 and 26°C the controller does nothing new; it keeps whatever it was doing. With a single threshold, the temperature would cross it, the fan would switch on, the temperature would fall a fraction of a degree, the fan would switch off, and the actuator would switch on and off in rapid succession around one value.

The size of the band sets the switching rate. Take an illustrative zone that, with the cooling off, warms toward 30°C with a time constant of $\tau = 10\,\mathrm{min}$, and with the cooling on cools toward 18°C with the same time constant. Within each interval of constant u, starting at time t_k, the temperature follows the first-order response of Chapter 8,

$$T(t) = T_{\infty,u} + \big(T(t_k) - T_{\infty,u}\big)\, e^{-(t-t_k)/\tau}, \qquad T_{\infty,0} = 30°\mathrm{C}, \quad T_{\infty,1} = 18°\mathrm{C}. \tag{10.8}$$

The off interval runs from 24 to 26°C and lasts $\tau \ln\big((30-24)/(30-26)\big) = 10 \ln 1.5 = 4.05\,\mathrm{min}$; the on interval runs from 26 down to 24°C and lasts $\tau \ln\big((26-18)/(24-18)\big) = 10 \ln(8/6) = 2.88\,\mathrm{min}$. One cycle takes 6.93 min, about 8.7 starts per hour. If the band is narrowed to 0.2 K, from 24.9 to 25.1°C, the same formula gives an off interval of $10 \ln(5.1/4.9)\,\mathrm{min} = 24\,\mathrm{s}$ and an on interval of $10 \ln(7.1/6.9)\,\mathrm{min} = 17\,\mathrm{s}$: a cycle of 41 s and about 87 starts per hour. Fans tolerate frequent switching better than compressors do, but every start costs wear, and compressors are commonly limited to a few starts per hour.

Figure 10.2 draws equation (10.8) for half an hour starting at 22°C with the cooling off. The upper panel shows the temperature in degrees Celsius against time in minutes, with the two thresholds as dashed lines; the lower panel shows the switching command u, 0 or 1, on the same time axis. The temperature rises to 26°C, the cooling switches on, the temperature falls to 24°C, the cooling switches off, and the pattern repeats as a saw-tooth confined to the band. The switching command changes only when the temperature touches a threshold. On/off control is simple and robust but holds the temperature only within a band, never at a value, and its cycling is the price of that simplicity.

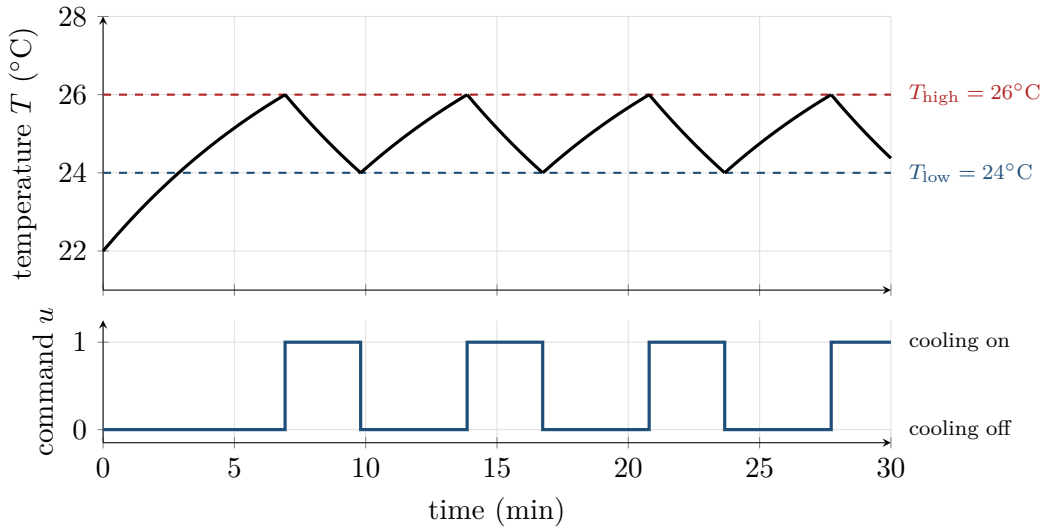


Figure 10.2: On/off control with a 2 K hysteresis band. Upper panel: zone temperature from equation (10.8) with $\tau = 10$ min, $T_{\infty,0} = 30°\text{C}$ (cooling off), $T_{\infty,1} = 18°\text{C}$ (cooling on), starting at 22°C; lower panel: the command of equation (10.7). Off intervals last 4.05 min and on intervals 2.88 min. Illustrative values.

10.2.3 Proportional control

A proportional controller makes the action grow with the size of the error instead of jumping between two values:

$$u = u_0 + K_p\, e, \qquad e = T - T_{\text{set}}, \tag{10.9}$$

where u_0 is the command at zero error and K_p the proportional gain. The law is often written $u = K_p e$; that is the same law with u measured as a deviation from u_0, and the rest of this section uses u in that sense. With an illustrative gain of $K_p = 10$ percentage points of fan speed per kelvin, a temperature 1 K above setpoint raises the fan speed slightly, by 10 points, and a temperature 5 K above setpoint raises it much more, by 50 points. The fan speed is bounded between a minimum and 100 %, so a large enough error drives the actuator into saturation, where the controller behaves like on/off control. Inside its range the proportional controller moves the actuator smoothly, without the cycling of Section 10.2.2.

Proportional control alone leaves a steady error when the load changes. To see why, write the zone as the linearized heat balance of Chapter 8 around an operating point at T_{set}: a thermal capacitance C_{th}, a conductance G through which the zone exchanges heat with its surroundings, a step $\Delta\dot{Q}$ in the heat load at $t = 0$, and an extra cooling ku for a command u above its operating value. Dividing by G,

$$\tau \frac{\mathrm{d}e}{\mathrm{d}t} = -e + \Delta - b\,u, \qquad \tau = \frac{C_{\text{th}}}{G}, \quad \Delta = \frac{\Delta\dot{Q}}{G}, \quad b = \frac{k}{G}, \tag{10.10}$$

where Δ is the temperature rise the load step would cause with no control and b the temperature reduction per unit of command. With illustrative values $\Delta\dot{Q} = 30$ kW, $G = 10$ kW/K, $k = 2$ kW per percentage point and $C_{\text{th}} = 6$ MJ/K, the plant has $\Delta = 3$ K, $b = 0.2$ K per point and $\tau = 6000/10 = 600$ s $= 10$ min. Without control the zone settles 3 K above setpoint, at 27°C. Substituting $u = K_p e$ into equation (10.10) and setting the derivative to zero gives

$$e_{\text{ss}} = \frac{\Delta}{1 + bK_p}, \tag{10.11}$$

where bK_p is the loop gain. With $K_p = 10$ points per kelvin the loop gain is $0.2 \times 10 = 2$, and $e_{\text{ss}} = 3/3 = 1$ K: the zone settles at 25°C against a setpoint of 24°C. The error is needed to

hold the extra fan speed, since a proportional controller only produces $u = 10$ points of extra fan speed if the error is 1 K. Raising the gain to 25 points per kelvin (loop gain 5) reduces the steady error to 0.5 K but cannot remove it, and in a real loop with lags (Section 10.2.6) a large gain causes oscillation.

10.2.4 Integral and proportional–integral control

An integral term removes the steady error. It accumulates the error over time,

$$u_I(t) = K_i \int_0^t e(s)\,\mathrm{d}s, \tag{10.12}$$

and keeps increasing the effort as long as any error of the same sign persists. The only state in which the integral stops changing is $e = 0$, so if the loop settles at all, it settles at the setpoint. Combined with the proportional term this gives the proportional–integral (PI) controller,

$$u = K_p\, e + K_i \int_0^t e(s)\,\mathrm{d}s, \tag{10.13}$$

which is the most common controller in heating, ventilation and air conditioning (HVAC): the proportional term reacts at once, the integral term trims away whatever error remains. In the plant of equation (10.10) the integral term ends holding $u = \Delta/b = 3/0.2 = 15$ points of extra fan speed, which at $k = 2$ kW per point is the 30 kW of extra cooling the load step requires. A controller with integral action must also stop integrating when its actuator saturates; otherwise the integral grows while the fan is already at full speed and the controller overshoots when the load falls. This protection is called anti-windup.

With the illustrative plant and $K_p = 10$ points per kelvin, choose the integral gain so that $bK_i = 0.225\,\mathrm{min}^{-1}$, that is $K_i = 1.125$ points per kelvin-minute. Writing $z = \int e\,\mathrm{d}t$, the closed loop is $\tau\ddot{z} + (1 + bK_p)\dot{z} + bK_i z = \Delta$, whose characteristic roots are both $s = -0.15\,\mathrm{min}^{-1}$ for these values (the critically damped case). The error is then

$$e_{\mathrm{PI}}(t) = \frac{\Delta}{\tau}\, t\, e^{-0.15\,t} = 0.3\, t\, e^{-0.15\,t}\ \mathrm{K}, \qquad t \text{ in minutes}, \tag{10.14}$$

which peaks at $t = 1/0.15 = 6.7$ min with $0.3 \times 6.67 \times e^{-1} = 0.74$ K and returns to zero. The proportional-only response is $e_{\mathrm{P}}(t) = 1 - e^{-0.3\,t}$ K, with closed-loop time constant $\tau/(1 + bK_p) = 3.3$ min, and the uncontrolled response is $3(1 - e^{-t/10})$ K.

Figure 10.3 plots the three responses after the load step. The horizontal axis is time in minutes; the vertical axis is the zone temperature $T_{\mathrm{set}} + e$. The dotted curve is the uncontrolled zone, which climbs to 27°C. The dashed curve is the proportional controller: it responds faster than the open zone and stops at 25°C, the steady error of equation (10.11). The solid curve is the PI controller: it rises to 24.7°C and then returns to the setpoint as the integral term builds up the extra 15 points of fan speed. The figure shows the division of labor in equation (10.13): the proportional term limits the excursion, the integral term removes the offset.

10.2.5 Proportional–integral–derivative control

The full PID law adds a term proportional to the rate of change of the error:

$$u = K_p\, e + K_i \int_0^t e(s)\,\mathrm{d}s + K_d\, \frac{\mathrm{d}e}{\mathrm{d}t}. \tag{10.15}$$

Each term looks at a different part of the error's history. The proportional term reacts to the present error. The integral term corrects a persistent error from the past. The derivative term anticipates the near future by reacting to the trend: if the temperature is 0.5 K above setpoint

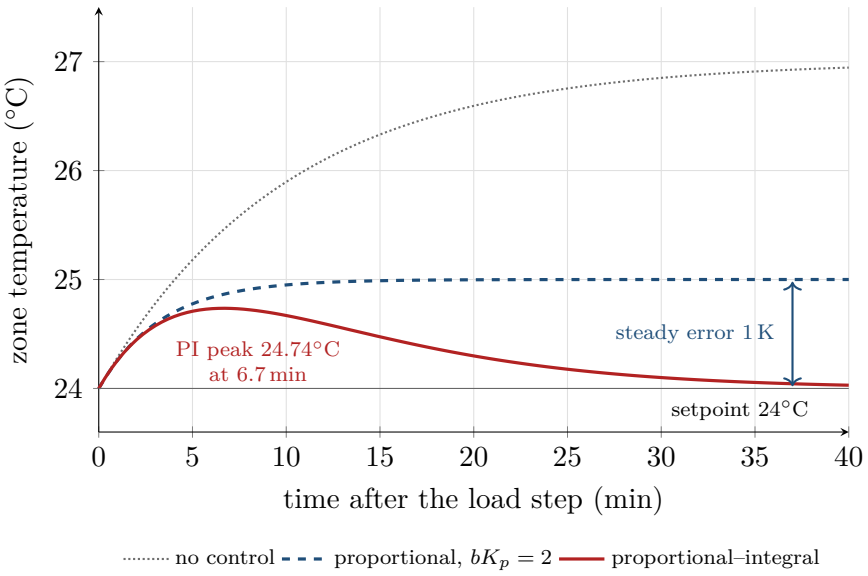


Figure 10.3: Response of the zone of equation (10.10) ($\tau = 10$ min, $\Delta = 3$ K, $b = 0.2$ K per point) to a 30 kW load step. No control: $24 + 3(1 - e^{-t/10})$. Proportional, $K_p = 10$ points/K: $25 - e^{-0.3t}$. PI, $K_i = 1.125$ points/(K min): $24 + 0.3\,t\,e^{-0.15t}$. Illustrative values.

and rising at 0.5 K per minute, a derivative gain of $K_d = 10$ points per (K/min) adds 5 points of fan speed now, before the error itself has grown. When the temperature is already falling back toward the setpoint, the derivative term is negative and eases the action, which damps overshoot.

The derivative term has a cost. It differentiates the measurement, and differentiation amplifies noise: a sensor with 0.1 K of random fluctuation sampled every 10 s produces apparent rates of change of the order of 0.1 K per 10 s, that is 0.6 K per minute, which with the gain above would move the fan by 6 points from noise alone. Derivative action therefore needs a filtered measurement, and in slow, noisy thermal loops it is often omitted. Not every loop needs all three terms: most HVAC loops run as PI, simple binary devices run as on/off, and derivative action is reserved for loops whose dynamics are slow compared with the sensor noise and whose overshoot must be kept small.

10.2.6 Tuning against thermal dynamics

The plant of a temperature loop is a heat balance,

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}}, \tag{10.16}$$

and its temperature responds slowly: a thermal capacitance must be charged or discharged before the temperature moves. Between the fan command and the temperature sensor there are further delays: the fan ramps up, the extra air takes time to travel to the rack, the sensor has its own time constant. An aggressive controller, one with a high gain or a fast integral, keeps raising the cooling while the temperature has not yet fallen, because it sees only the error that persists. By the time the temperature responds, the cooling is far larger than needed, the temperature falls below the setpoint, the controller backs off too far, and the temperature rises again: the loop oscillates.

An example of such a loop is a temperature that moves $24 \to 27 \to 21 \to 26 \to 22 \to 25$°C while the fan speed swings $40 \to 100 \to 20 \to 90 \to 30$ %. Figure 10.4 draws that sequence against a well-tuned response to the same load step. The upper panel is temperature against

time in minutes and the lower panel fan speed in percent on the same time axis. In the badly tuned loop (solid) each fan extreme follows the corresponding temperature extreme by about 2.5 min, the swings decay slowly, and the fan travels across most of its range every few minutes. In the well-tuned loop (dashed) the temperature rises about 1 K and returns, and the fan moves once to its new level. Nothing is broken in the oscillating loop: the fans, coil and sensor all work, and the same equipment with different gains gives the dashed response. That is a diagnostic distinction worth keeping. A regular oscillation in which the actuator swings in phase-lagged step with the temperature points to tuning, not to a failed component; a fault, such as a lost fan or a blocked filter, appears instead as a shift in the level of temperature or cooling that the controller cannot remove.

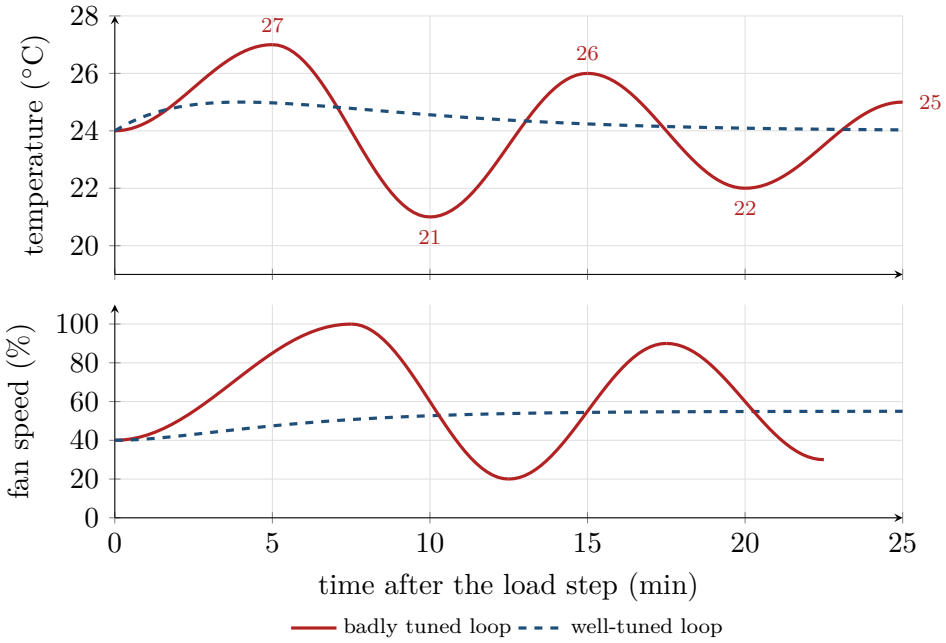


Figure 10.4: A badly tuned loop against a well-tuned one after the same load step. Solid: the example sequence $24, 27, 21, 26, 22, 25\,^{\circ}\mathrm{C}$ and $40, 100, 20, 90, 30\,\%$, joined by half-cosine segments between the stated extrema (temperature extrema every 5 min, fan extrema 2.5 min later). Dashed: $24 + (t/4)\,e^{1-t/4}$ and $40 + 15(1 - (1 + t/3)e^{-t/3})$. Illustrative curves.

Principle 10.2. *A thermal plant responds slowly, so a controller that acts on the persistence of the error overshoots; oscillation with healthy equipment is a tuning fault, and it must be told apart from an equipment fault before anything is repaired.*

10.3 The loops of a cooling plant

A cooling plant is a set of feedback loops of the kind just described, nested and coupled. This section follows them from the air side to the chillers: the cascade that controls rack temperature through airflow, the loops that hold the chilled-water temperature, the coil valves and the loop pressure, and the discrete decision of how many chillers to run.

10.3.1 Cascaded control

Many temperature loops are built as a cascade of two controllers. The outer loop measures the rack temperature and computes the airflow needed to hold it, $T_{\mathrm{rack}} \rightarrow \dot{V}_{\mathrm{set}}$. The inner loop

measures the airflow and adjusts the fan speed to deliver it, $\dot{V}_{\text{set}} \rightarrow N_{\text{fan}}$. The outer controller does not command the fan directly; it commands an airflow and leaves the fan to the inner controller.

Figure 10.5 draws the structure. The forward path runs left to right through the two comparisons and the two controllers, then through the fan and the rack. The short feedback path below the inner part returns the measured airflow to the inner comparison; the long feedback path at the bottom returns the measured rack temperature to the outer comparison. Both comparisons use measurement minus setpoint, the convention of Section 10.2.1. The cascade has two advantages. A disturbance inside the inner loop, such as a filter that loads up and reduces the airflow delivered at a given fan speed, is detected by the airflow measurement and corrected within seconds, before it reaches the rack temperature. And the outer loop sees a well-behaved inner loop, an airflow that follows its setpoint, instead of the nonlinear relation between fan speed and airflow. The structure works only if the inner loop is several times faster than the outer one, a ratio of five to ten in time constants being a common rule; otherwise the two controllers act on the same time scale and interfere.

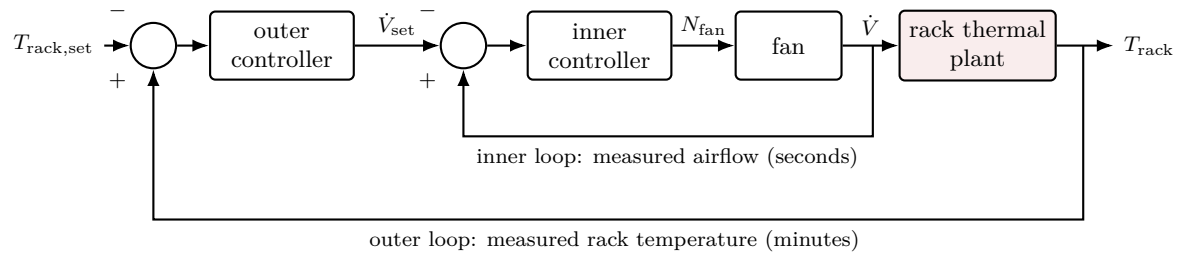


Figure 10.5: Cascaded control of rack temperature. The outer controller turns the rack temperature error into an airflow setpoint; the inner controller turns the airflow error into a fan speed. The inner loop corrects airflow disturbances before they reach the rack temperature.

10.3.2 Chilled-water supply temperature

On the water side the main loop holds the chilled-water supply temperature $T_{\text{CHW,s}}$ at its setpoint $T_{\text{CHW,set}}$, for example 7°C, by changing the loading of the chiller compressor. The chiller removes heat from the returning water in its evaporator,

$$\dot{Q}_{\text{evap}} = \dot{m}_w \, c_{p,w} \, (T_{\text{CHW,r}} - T_{\text{CHW,s}}), \tag{10.17}$$

and the compressor provides the work that moves that heat to the condenser. If the supply temperature drifts above 7°C, because the load has grown or the return water has become warmer, the controller raises the compressor loading so that more heat is removed and the supply temperature returns to its setpoint; the compressor works harder and draws more power. If the supply temperature drops below 7°C, the controller unloads the compressor.

With illustrative numbers, a load of 1.5 MW carried with a 5 K difference between return and supply needs a water flow of $\dot{m}_w = 1500/(4.18 \times 5) = 71.8$ kg/s. The supply loop does not control that flow; the pumps and valves do. It controls only the temperature, and it therefore treats the flow and the return temperature as disturbances. The supply setpoint itself is a control input of the facility, the fourth entry of equation (10.3), and Section 10.4.5 returns to the choice of its value.

10.3.3 Coil valve and pump pressure

Two further loops share the chilled water among the air handlers.

Coil valve. At each computer room air handler (CRAH) a valve controls the chilled-water flow through the cooling coil. If the supply air temperature leaving the unit is above its setpoint, $T_{\text{air,supply}} > T_{\text{air,set}}$, the valve opens: the water flow $\dot{m}_w$ through the coil rises, the heat the coil removes $\dot{Q}_{\text{coil}}$ rises, and the supply air temperature falls,

$$T_{\text{air,supply}} > T_{\text{air,set}} \implies u_{\text{valve}} \uparrow \implies \dot{m}_w \uparrow \implies \dot{Q}_{\text{coil}} \uparrow \implies T_{\text{air,supply}} \downarrow .$$

This is another negative feedback loop, one per air handler. Its gain is not constant: a coil's heat transfer rises steeply with flow at low flow and flattens at high flow, so a valve that opens from 10 to 20 % changes the cooling much more than one that opens from 80 to 90 %. That nonlinearity is one reason the valve loop is commonly placed inside a cascade, as in Section 10.3.1.

Pump differential pressure. The valves can deliver flow only if there is pressure across them. The pump loop holds a differential pressure setpoint Δp_{set}, usually measured across the supply and return headers near the most remote coil. When remote valves open, the loop resistance falls, the differential pressure drops below its setpoint, and the pump speed rises to restore it; when valves close, the pressure rises and the pump slows. The pump loop thus makes sure that many coils, each with its own valve loop, receive adequate flow. The two loops are coupled: every valve movement is a disturbance to the pump loop, and every pump speed change is a disturbance to all the valve loops. Because pump power grows roughly with the cube of the speed, the choice of Δp_{set} matters for energy: a setpoint higher than the most demanding valve needs makes the pump run faster than necessary, and the valves throttle away the excess pressure.

10.3.4 Chiller staging

The number of chillers running is a discrete control input. Suppose each chiller provides 2 MW of cooling. At a demand of 1.5 MW one chiller is enough; at 3.2 MW one chiller cannot carry the load and a second must start. The rule has the form

$$\dot{Q}_{\text{load}} > \dot{Q}_{\text{threshold}}(N_{\text{ch}}) \implies N_{\text{ch}} \leftarrow N_{\text{ch}} + 1, \tag{10.18}$$

with a matching rule for removing a chiller when the load falls. The decision is discrete because a chiller is either running or not, and because each chiller has a minimum stable loading below which it cannot run continuously.

Staging rules use the same device as on/off control: a hysteresis band. An illustrative rule starts an additional chiller when the load exceeds 90 % of the running capacity, $\dot{Q}_{\text{load}} > 0.9 \times 2N_{\text{ch}}$ MW, and stops one when the load falls below 70 % of the capacity that would remain, $\dot{Q}_{\text{load}} < 0.7 \times 2(N_{\text{ch}} - 1)$ MW. A second chiller then starts at 1.8 MW and stops at 1.4 MW; a third starts at 3.6 MW and stops at 2.8 MW; a fourth starts at 5.4 MW and stops at 4.2 MW. Figure 10.6 plots the number of chillers running against the load. The horizontal axis is the cooling load in megawatts and the vertical axis the number of chillers. The solid staircase is the path followed while the load rises, the dashed staircase the path followed while it falls, and the dotted staircase the minimum number whose capacity covers the load. Between the two paths lies a band in which the number running depends on the history: at 1.5 MW the plant runs one chiller if the load has been rising and two if it has been falling from above 1.8 MW. At 3.2 MW a rising load runs two chillers, the second having been started at 1.8 MW, before the first chiller reached its limit.

10.3.5 Staging as a trade-off

Staging is not a simple threshold decision because both ways of getting it wrong cost something. Starting a chiller costs energy: the compressor, its condenser-water pump and its tower fan

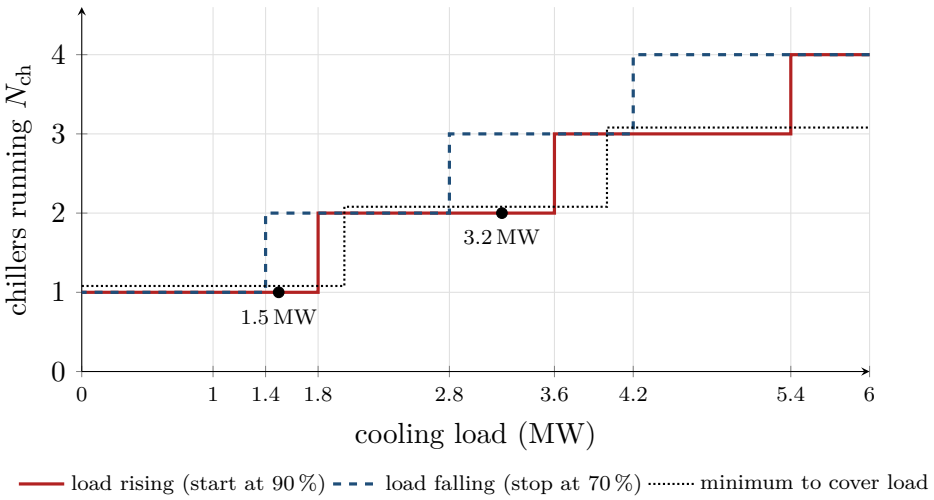


Figure 10.6: Chiller staging against load for 2 MW chillers with an illustrative hysteresis rule: start an additional chiller above 90 % of running capacity, stop one below 70 % of the capacity that would remain. The dotted staircase, $\lceil \dot{Q}_{\text{load}}/2\,\text{MW} \rceil$, is drawn slightly raised for visibility. Dots: the two example loads on a rising load.

start drawing power, and two lightly loaded chillers can consume more than one well loaded chiller. Each start also adds wear, and a chiller that starts and stops within minutes has been run inefficiently for no benefit. Waiting too long has the opposite cost: the running chillers approach their limit, the plant has no spare capacity if the load rises further or a unit trips, and the chilled-water temperature may drift up before the new chiller has reached full output.

The trade-off is between energy efficiency and capacity margin, and the hysteresis band of Figure 10.6 is a fixed compromise between them. A better compromise depends on what the load will do next. If demand is about to rise, starting the next chiller early costs a few minutes of inefficient operation and protects the margin; if demand is about to fall, the same start is wasted. A controller that knows only the present load cannot tell the two situations apart. A controller that has a forecast of the load can, and the next section begins there.

10.4 Predictive and coordinated control

The loops of the previous section react to what has already happened. This section adds what is expected to happen: forecasts, model predictive control with its cost function and constraints, and the coordination of loops whose individual optima do not add up to the best operation of the facility.

10.4.1 Forecast-assisted control

Consider a facility carrying a cooling load of 8 MW in a plant configuration that supports 9 MW. A load forecast says

$$\dot{Q}_{\text{load}}(t + 20\,\text{min}) = 10.5\,\text{MW}. \tag{10.19}$$

A reactive controller does nothing until the load reaches the configuration's limit. A predictive controller starts equipment now, so that the additional capacity is running before it is needed.

Assume, for illustration, that the load rises linearly from 8 to 10.5 MW over the 20 minutes, that is 0.125 MW/min, and that an additional chiller needs 6 minutes from the start command to full output. The load reaches 9 MW after $(9 - 8)/0.125 = 8$ min. The reactive controller starts the chiller at that moment and has it at full output at 14 min; between 8 and 14 min the

load exceeds the running capacity by up to $8 + 0.125 \times 14 - 9 = 0.75\,\mathrm{MW}$, and that deficit is carried by the thermal mass of the water loop and the room, whose temperatures rise. The predictive controller starts the chiller when the forecast arrives, at $t = 0$, and has it at full output at 6 min, when the load is 8.75 MW, still inside the original capacity. The chain is forecast, then anticipatory action, and the value of the forecast lies entirely in the time it gives the plant to respond. That value depends on the forecast's reliability: a forecast that is often wrong starts chillers for nothing, and the decision of how much to trust it is a question of uncertainty that Section 10.5 takes up.

10.4.2 Model predictive control

Model predictive control turns forecast-assisted control into a procedure. At each time t it

1. estimates the current state $\hat{x}_t$;
2. predicts the future states with the model, using forecasts of the disturbances;
3. tests candidate control actions against those predictions;
4. chooses the best sequence of actions over a horizon of H steps;
5. applies only the first action of that sequence;
6. moves to $t + 1$ and repeats.

Written as an optimization, step 4 is

$$\begin{aligned} &\min_{u_t,\dots,u_{t+H-1}} \ \sum_{k=t}^{t+H-1} c(x_k, u_k) \\ &\text{subject to} \quad x_{k+1} = f(x_k, u_k, d_k), \quad x_t = \hat{x}_t, \quad T_{\mathrm{rack},i,k} \le T_{\max} \text{ for all } i, k, \end{aligned} \tag{10.20}$$

where f now denotes the one-step map of a discretized version of the model of equation (10.1), d_k the forecast disturbance at step k, c the cost of one step, and i indexes the racks.

Figure 10.7 draws one solve. Both panels share a time axis in control steps, with the present time t at zero and the shaded region marking the horizon of $H = 8$ steps. The upper panel shows the rack temperature: the solid curve to the left of t is the measured past; to the right, the dotted curve is the prediction if the command stays at its present value, which crosses the limit $T_{\max}$ (dashed line) within four steps, and the dashed curve is the prediction under the optimized plan, which stays below the limit. The lower panel shows the command, a fan speed: constant in the past, then the planned sequence over the horizon, of which only the first step, filled, is applied. At $t + 1$ the state is measured again, the horizon slides forward by one step, and the remaining seven planned steps are discarded and recomputed with the new information. The repetition is what makes the horizon recede, and it is also what makes the scheme a feedback controller: a prediction error at one step is corrected at the next.

10.4.3 Cost function and constraints

The stage cost c in equation (10.20) states what the facility wants. A typical form has three terms,

$$c(x, u) = w_1\, P_{\mathrm{cool}} + w_2\, (T - T_{\mathrm{set}})^2 + w_3\, (\Delta u)^2, \tag{10.21}$$

where P_{cool} is the electrical power of the cooling system, $T - T_{\mathrm{set}}$ the deviation of a controlled temperature from its setpoint, $\Delta u = u_k - u_{k-1}$ the change in the command from one step to the next, and w_1, w_2, w_3 non-negative weights. The first term penalizes energy, the second penalizes temperature deviation, and the third penalizes aggressive control changes, which wear out equipment and excite the oscillations of Section 10.2.6. The weights convert the three into one currency and express the operator's priorities: with a large w_3 the plan of Figure 10.7 changes the fan speed in smaller steps and accepts a slower correction.

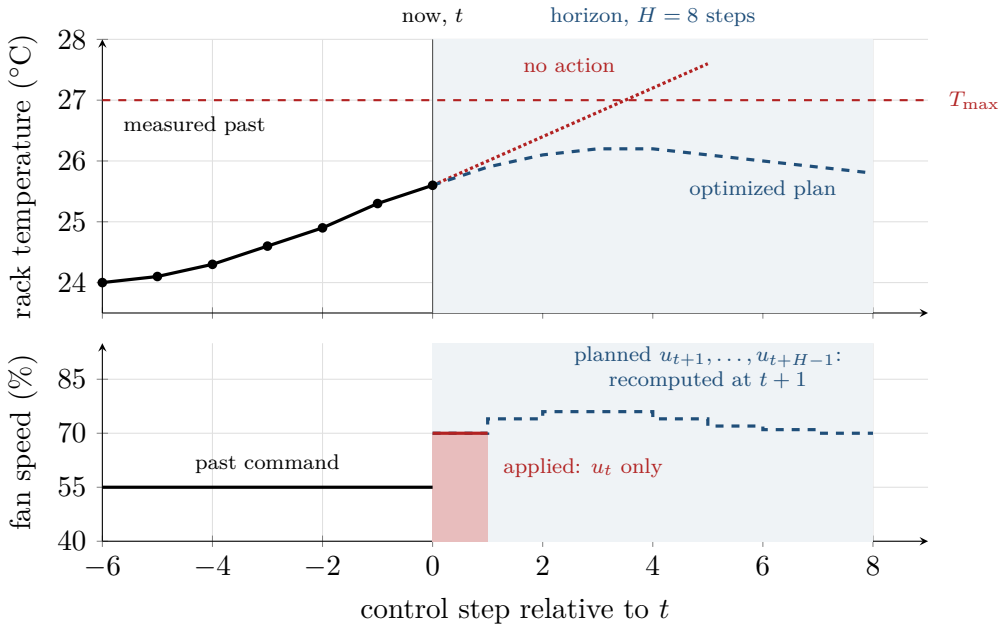


Figure 10.7: One solve of model predictive control, equation (10.20), with a horizon of $H = 8$ steps (shaded). Upper panel: measured rack temperature up to t, predicted temperature with the command held at its present value (crosses $T_{\max}$) and under the optimized plan. Lower panel: the planned fan-speed sequence; only the first step is applied before the problem is solved again at $t + 1$. Illustrative values.

The constraints matter more than the objective. A cooling controller is allowed to trade energy against comfort only inside a region of safe operation, and that region is written as hard limits:

$$T_{\text{rack},i} \le T_{\max}, \qquad \dot{m}_w \le \dot{m}_{w,\max}, \qquad 0 \le u_{\text{valve}} \le 1, \qquad N_{\text{fan,min}} \le N_{\text{fan}} \le N_{\text{fan,max}}. \tag{10.22}$$

The first is a safety limit on every rack, the second a limit of the pumps and pipes, the third and fourth the physical range of the actuators. A weighted temperature term in equation (10.21) would let a large energy saving buy a small excursion above $T_{\max}$; a constraint does not. The problem a data center controller solves is therefore to minimize energy subject to safety constraints, not to balance energy against safety.

Principle 10.3. *In a cooling optimizer the safety limits are constraints, not terms of the cost: the objective chooses among safe operating points and never buys an unsafe one.*

10.4.4 Coordination of loops

Loops that are each correct can work against each other. An air handler controller raises its fan speed because a rack temperature is high. At the same time a plant efficiency controller, seeing a light load on the chillers, lowers the chilled-water flow to save pump energy. The reduced water flow makes the coil supply warmer air, the rack temperature rises further, and the fan controller pushes harder; the fans now spend the energy the pumps saved, and more. Each controller is doing what it was designed to do, and the facility as a whole is worse off. The remedy is coordination at the level of the whole system: a controller, or a set of rules, that sees both decisions and their common effect on the rack temperature.

Local versus global optimum. The same structure appears in setpoints. Each air handler may hold a supply air temperature of 18°C correctly and steadily, while the facility could safely supply 22°C air, with every rack inlet still below its limit, and use much less chiller energy (Section 10.4.5 quantifies the saving). Each unit's control is locally correct; the setpoint it is given is globally suboptimal. Control that is correct for every loop is not the same as control that is optimal for the facility, and the second requires a model of how the loops interact.

10.4.5 Setpoints and the chilled-water temperature

The low-level loops of Section 10.3 hold setpoints; something above them chooses the setpoints. The main ones are the supply air temperature $T_{\mathrm{air,set}}$, the chilled-water supply temperature $T_{\mathrm{CHW,set}}$, the loop differential pressure Δp_{set} and the humidity setpoint $\mathrm{RH}_{\mathrm{set}}$. Many facilities run with conservative historical setpoints, such as 18°C supply air where 22°C is safe, chosen once at commissioning with a wide margin and never revisited. Each such setpoint costs energy every hour, and setpoint optimization, the choice of setpoints from the present state and load, is one of the cheapest sources of savings because it changes no equipment.

The chilled-water supply temperature shows the trade-off. Raising it reduces the lift of the chiller, the temperature difference between the evaporator and the condenser, and a smaller lift gives a higher coefficient of performance. As an illustration, take an evaporating temperature 2 K below the chilled-water supply and a condensing temperature of 35°C (308.15 K). At a supply of 7°C the evaporator is at 5°C (278.15 K) and the ideal (Carnot) COP is $278.15/30 = 9.27$; at a supply of 11°C the evaporator is at 9°C and the ideal COP is $282.15/26 = 10.85$, 17 % higher. If the real chiller achieves a fixed fraction of 0.55 of the ideal value, its COP rises from 5.10 to 5.97, and a 1.5 MW load costs 294 kW of compressor power at 7°C and 251 kW at 11°C, a saving of 43 kW. The cost is margin: warmer chilled water gives warmer supply air and warmer rack inlets. The decision is therefore

$$\max\ T_{\mathrm{CHW,supply}} \quad \text{subject to} \quad T_{\mathrm{rack},i} \le T_{\max} \text{ for all racks } i, \tag{10.23}$$

and the binding constraint is set by the worst rack, the one with the least margin, not by the average.

10.4.6 Fan energy against thermal margin

Fan power follows the fan affinity laws: airflow is proportional to speed and power to the cube of speed, $P_{\mathrm{fan}} \propto N_{\mathrm{fan}}^3$. A 10 % reduction in speed reduces fan power to $0.9^3 = 0.729$ of its value, a 27 % saving; a 20 % reduction reduces it to 0.512. The same reduction lowers the airflow proportionally, and the air temperature rise across the IT equipment grows as $\Delta T = \dot{Q}_{\mathrm{rack}}/(\dot{m}_{\mathrm{air}} c_{p,\mathrm{air}})$. For an illustrative 20 kW rack with 1.6 kg/s of air, the rise is $20/(1.6 \times 1.0) = 12.5$ K; with 10 % less air it is 13.9 K. As long as the air handlers supply at least as much air as the racks draw, the inlet temperatures change little; once they supply less, hot exhaust is drawn back into the inlets, the recirculation fraction rises, and the rack inlet temperatures climb (Chapter 6).

The chain is $N_{\mathrm{fan}} \downarrow \Rightarrow \dot{m}_{\mathrm{air}} \downarrow \Rightarrow T_{\mathrm{rack}} \uparrow$, and it sets a trade-off between efficiency and thermal risk. The cubic law rewards every reduction in speed; the airflow balance punishes the reduction that goes one step too far. The efficient operating point lies close to the boundary at which the supplied airflow just covers the demand, and the controller's task is to operate near that boundary without crossing it. How close it can go depends on how well it knows where the boundary is, which is a question of the uncertainty of Section 10.5.

10.4.7 Redundancy and unequal efficiency

Redundancy adds a control layer of its own. A plant with four chillers of which three are needed (an $N{+}1$ arrangement) must decide which chillers run, which one is the spare, and how the load is shared among those running. Equal sharing is the simplest rule and the usual default, but it is not optimal when the chillers differ.

Chillers do differ. Suppose three of them have rated COPs of $\mathrm{COP}_1 = 6$, $\mathrm{COP}_2 = 5$ and $\mathrm{COP}_3 = 4$, because of age, fouling or design. Their COP also depends on their loading ℓ, the fraction of capacity in use: a chiller is least efficient at very low loading and loses efficiency again near full capacity. With an illustrative part-load factor $\mathrm{COP}_j(\ell) = \mathrm{COP}_j\,(1 - 1.5(\ell - 0.6)^2)$ and 2 MW chillers, Table 10.2 compares two dispatches at two loads; each row gives the chillers' loads, their loadings and COPs, and the total compressor power $\sum_j \dot{Q}_j/\mathrm{COP}_j(\ell_j)$. At a low load of 0.8 MW the most efficient chiller alone needs 142 kW, while splitting the load equally between chillers 1 and 2 pushes both to a loading of 0.2 and needs 193 kW: one chiller wins. At 1.9 MW chiller 1 alone runs at 95 % loading with a COP of 4.90 and needs 388 kW, while chillers 1 and 2 sharing 1.1 and 0.8 MW need 354 kW: two chillers beat one near full capacity, even though the second chiller is less efficient. The least efficient chiller, with $\mathrm{COP}_3 = 4$, is the natural spare.

Table 10.2: Chiller dispatch with unequal efficiency: compressor power for one and two running chillers at two loads. Part-load factor $1 - 1.5(\ell - 0.6)^2$, 2 MW chillers; auxiliary pump and tower power is not included. Illustrative values.

Load	Dispatch	Loading ℓ	COP	Power (kW)
0.8 MW	chiller 1 alone	0.40	5.64	142
0.8 MW	chillers 1 and 2, $0.4 + 0.4$ MW	0.20, 0.20	4.56, 3.80	193
1.9 MW	chiller 1 alone	0.95	4.90	388
1.9 MW	chillers 1 and 2, $1.1 + 0.8$ MW	0.55, 0.40	5.98, 4.70	354

The decision of which units to run and at what output, given each unit's efficiency curve, start-up cost and minimum loading, has the same structure as unit commitment in a power system, where a system operator decides which generators to commit and how to dispatch them to meet a load at least cost. It combines a discrete choice (which chillers are on) with a continuous one (how the load is shared), and the staging rule of Section 10.3.4 is its simplest heuristic.

10.5 Control under uncertainty

Every controller so far acted as if the state were known. This section removes that assumption: the state is a probability distribution, a constraint becomes a probability, and the controller is either robust to a range of parameter values or optimal on average over them; in every case the state must be estimated before it can be controlled.

10.5.1 The uncertain state

A controller does not see the state; it sees measurements. What it can know about the state is a probability distribution conditioned on the measurements,

$$x \sim p(x \mid y), \tag{10.24}$$

and many of the quantities that matter for control are uncertain under that distribution: the airflow through each rack, which is rarely measured; the heat-transfer coefficient of each coil, which drifts with fouling; the heat load of each rack, which is often estimated from power-distribution readings at a coarser level; the bias of each sensor; and the capacity of each piece of equipment, which may be below its nameplate.

The measured, calculated and inferred distinction applies here directly. A rack inlet temperature is measured, with a known noise and an unknown bias. A coil heat removal is calculated from measured water flow and temperatures, and inherits their errors. A rack airflow, a recirculation fraction or a coil conductance is inferred, and its uncertainty is whatever the measurements leave unresolved. A controller that treats an inferred quantity as if it were measured acts with more confidence than the data support.

10.5.2 The chance constraint

A deterministic constraint requires a rack temperature to stay below a limit,

$$T_{\text{rack}} < 30^{\circ}\text{C}. \tag{10.25}$$

When the rack temperature is known only as a distribution, the constraint can be written as a probability instead,

$$\Pr(T_{\text{rack}} > 30^{\circ}\text{C}) < 0.01, \tag{10.26}$$

a chance constraint. It says that the controller must keep the probability of exceeding the limit below one in a hundred. The level 0.01 is a risk level the operator chooses: the controller no longer promises that the limit will never be exceeded, which no controller working from uncertain measurements can honestly promise; it holds the facility at a specified risk.

For a normally distributed prediction with mean μ and standard deviation σ, the chance constraint reduces to a deterministic constraint with a margin,

$$\Pr(T > T_{\max}) < \epsilon \quad \Longleftrightarrow \quad \mu + z_{1-\epsilon}\,\sigma < T_{\max}, \tag{10.27}$$

where $z_{1-\epsilon}$ is the standard normal quantile, 2.326 for $\epsilon = 0.01$. The margin is proportional to the uncertainty: a better estimate of the state, a smaller σ, lets the controller operate closer to the limit at the same risk.

10.5.3 Same mean, different risk

Consider two racks whose predicted inlet temperatures both have mean 27°C. The first is well instrumented and its prediction has a standard deviation of 0.5 K; the second depends on an inferred airflow and its prediction has a standard deviation of 4 K. A deterministic controller looks only at the means and treats the two racks alike, both 3 K below a 30°C limit. Their risks are very different. For the first rack the limit is $3/0.5 = 6$ standard deviations above the mean, and $\Pr(T > 30^{\circ}\text{C}) = 1 - \Phi(6) \approx 1 \times 10^{-9}$. For the second it is $3/4 = 0.75$ standard deviations above, and $\Pr(T > 30^{\circ}\text{C}) = 1 - \Phi(0.75) = 0.227$: almost one chance in four. Against the chance constraint of equation (10.26), the first rack satisfies it with a wide margin, and the second violates it by a factor of more than twenty. By equation (10.27), the first rack may run with a mean up to $30 - 2.326 \times 0.5 = 28.8^{\circ}\text{C}$; the second must be brought down to $30 - 2.326 \times 4 = 20.7^{\circ}\text{C}$, or its uncertainty must be reduced.

Figure 10.8 draws the two predictions. The horizontal axis is the rack inlet temperature and the vertical axis the probability density. The tall narrow curve is the first rack and the low wide curve the second; the dashed vertical line is the 30°C limit, and the shaded area under the wide curve to the right of the limit is its probability of exceeding it, 0.227. The narrow curve has no visible area beyond the limit. The figure shows why a controller that works with means alone is blind to the difference: the two curves are centered at the same point.

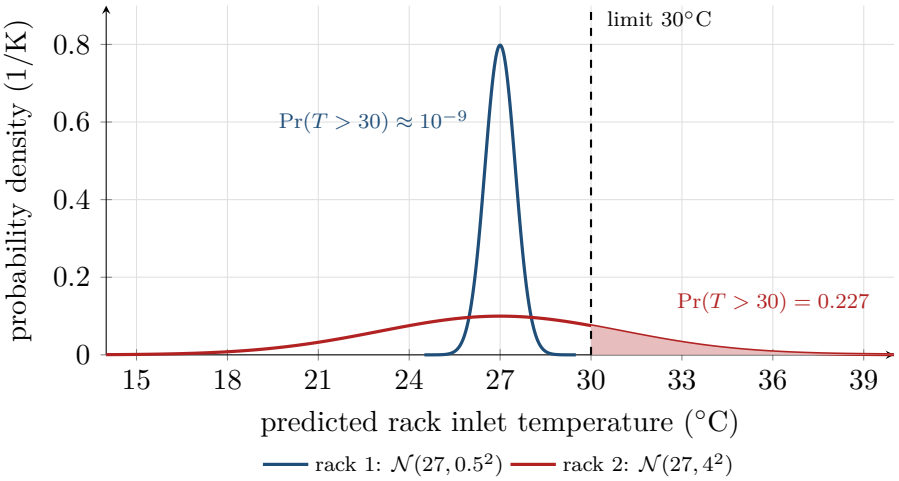


Figure 10.8: Two predicted rack inlet temperatures with the same mean, 27°C, and standard deviations 0.5 K and 4 K (normal densities). The shaded tail beyond the 30°C limit is $1 - \Phi(0.75) = 0.227$ for the wide prediction; for the narrow one it is $1 - \Phi(6) \approx 10^{-9}$. Illustrative values.

Principle 10.4. *A controller should act on the risk of crossing a limit, not on the predicted value alone: two racks with the same predicted temperature can carry risks that differ by eight orders of magnitude.*

10.5.4 Robust and stochastic control

There are two standard ways to choose a control when parameters are uncertain. Robust control declares a set of plausible values and requires the control to work for all of them. For a thermal resistance known only to lie in an interval,

$$R_{\text{th}} \in [R_{\min}, R_{\max}], \tag{10.28}$$

the robust controller chooses u such that every rack constraint holds for every R_{th} in the interval. Take an illustrative rack whose inlet temperature is $T_{\text{rack}} = T_{\text{supply}} + R_{\text{th}}\,\dot{Q}_{\text{rack}}$, with $\dot{Q}_{\text{rack}} = 10\,\text{kW}$, $R_{\text{th}} \in [0.6, 1.0]\,\text{K/kW}$ (the resistance standing for the mixing between the supply air and the inlet) and $T_{\max} = 32°\text{C}$. The robust supply temperature must satisfy $T_{\text{supply}} + 1.0 \times 10 \le 32$, that is $T_{\text{supply}} \le 22°\text{C}$, while at the central value 0.8 K/kW the limit would be 24°C. Robust control is safe by construction and often conservative, because the worst case at the edge of the interval may be far less likely than the central value.

Stochastic control instead treats the parameters as random, $\theta \sim p(\theta)$, and optimizes the expected cost,

$$\min_u \ \mathbb{E}_\theta\big[c(x,u)\big] \qquad \text{or} \qquad \min_u \ \mathbb{E}_\theta\big[c(x,u)\big] + \lambda\,\text{Risk}(x,u), \tag{10.29}$$

where Risk is a measure of the bad outcomes, for example the probability of exceeding a limit or the expected size of the exceedance, and $\lambda \ge 0$ sets how much expected cost the operator is willing to pay to reduce it. The chance constraint of Section 10.5.2 is a third form, in which the risk is held below a level rather than priced. Robust control uses only the range of the parameters; stochastic control and the chance constraint use their whole distribution and can therefore be less conservative where the distribution is known.

10.5.5 Estimation before control

All three forms need the distribution of the state and the parameters, and that distribution comes from state estimation. The full chain is

$$\begin{aligned}\text{physical system} \to \text{sensors} \to \text{state estimation} \to \text{optimization and control}\\ \to \text{actuators} \to \text{physical system.}\end{aligned}$$

The estimation step turns noisy, biased and incomplete measurements into an estimate of the state and its uncertainty, $p(x \mid y)$; the control step chooses u from that estimate. A controller is only as good as the state it acts on: a well-tuned PI loop holding a biased sensor at its setpoint holds the real temperature at the wrong value, and an optimizer that pushes the chilled-water temperature to the limit of equation (10.23) pushes it to the wrong limit if the worst rack's temperature is mis-estimated.

The order matters for the uncertainty as well. The margins of equation (10.27) scale with the standard deviation of the estimate, so the energy a facility can save by operating close to its limits is bounded by how well it knows its state. A better estimate is, in that sense, a control action: it moves the safe operating point toward the limit without changing any equipment.

10.6 Diagnosis-aware and supervisory control

This section places the controller in its operating context: actions that depend on what is believed to be wrong, a simple safety layer beneath a sophisticated optimizer, the IT load as a control input, the response to a chiller failure, and the supervisory layer that ranks actions by the risk they remove.

10.6.1 Diagnosis-aware action

A rack temperature rises. A naive controller raises the air handler fan speed, because that is what its loop does for any positive error. Whether that helps depends on the cause. If the cause is a sensor bias, the real temperature is normal and the extra fan speed wastes energy while the sensor still reads high. If the cause is a blocked airflow path to that rack, a faster air handler fan does little, because the air it adds does not reach the rack. The action should therefore depend on the probabilities of the hypotheses given the measurements, $\Pr(H_i \mid y)$, not on the temperature alone.

Table 10.3 gives an example. Three hypotheses explain the temperature rise, with posterior probabilities 0.65 for low airflow from the air handler, 0.25 for a rack sensor bias and 0.10 for a high IT load. The table lists, for each hypothesis, what a fan increase would achieve and what the appropriate action is. No single action is right under all three. The combined response weighted by the probabilities is to verify the air handler's fan status, raise the airflow near the rack cautiously, and request a second temperature measurement.

The second measurement deserves a closer look, because it may be the best action of all. Suppose, illustratively, that a second sensor at the same rack reads high with probability 0.95 if the rack is really hot (H_1 or H_3) and with probability 0.05 if the first sensor is biased (H_2). The second sensor reads high with probability $0.65 \times 0.95 + 0.25 \times 0.05 + 0.10 \times 0.95 = 0.725$, and Bayes' rule then gives

$$\begin{aligned}\text{second sensor reads high:}&\quad (\Pr(H_1 \mid y),\ \Pr(H_2 \mid y),\ \Pr(H_3 \mid y)) = (0.852,\ 0.017,\ 0.131),\\ \text{second sensor reads normal:}&\quad (\Pr(H_1 \mid y),\ \Pr(H_2 \mid y),\ \Pr(H_3 \mid y)) = (0.118,\ 0.864,\ 0.018).\end{aligned}$$

One reading moves the sensor-bias hypothesis from 0.25 to either 0.02 or 0.86. In the first case the facility acts on the airflow with confidence; in the second it repairs the sensor and spends no cooling energy. The best action may be to measure first: control and the value of information

Table 10.3: A diagnosis-aware response to a rising rack temperature: three hypotheses, their probabilities, and what each implies for action. Illustrative values.

Hypothesis H_i	$\Pr(H_i \mid y)$	Effect of raising fan speed	Appropriate action
H_1: low CRAH airflow	0.65	helps if the fan has slowed; little if the path is blocked	verify CRAH fan status; raise nearby airflow cautiously
H_2: rack sensor bias	0.25	none on the real temperature; wastes energy	take a second, independent temperature measurement
H_3: high IT load	0.10	helps, at an energy cost	confirm rack power; raise cooling or cap the load

meet, and a measurement is worth taking when the decision it may change is worth more than the delay and the cost of taking it.

10.6.2 Safety layer and optimization layer

A practical control architecture separates two layers. The safety layer is fast, conservative and simple: a rule such as

$$T > T_{\text{crit}} \implies \text{maximum cooling}, \tag{10.30}$$

which needs no model, no forecast and no optimization, and which acts within seconds of the measurement. The optimization layer is slower and sophisticated: it minimizes P_{cool} with safe margins, using estimates, forecasts and models, and it acts on the time scale of minutes by changing setpoints. The optimizer may be as elaborate as the models allow; the safety layer must stay simple and reliable, because it is what protects the equipment when the optimizer is wrong, when its model is out of date, or when its inputs fail.

Figure 10.9 draws the architecture. The boxes in the top row are the supervisory layer, which runs from state estimation through diagnosis and risk to an optimizer that chooses setpoints. The right column carries those setpoints down to the low-level PID loops of Section 10.2 and Section 10.3, which drive the actuators. The bottom box is the physical system; its measurements rise through the sensors on the left back to the estimator. The safety layer in the middle row takes the sensor readings directly and can override the actuators (red dashed arrow) without passing through the estimator or the optimizer. The figure shows the separation of time scales: the supervisory loop closes in minutes, the low-level loops in seconds, and the safety path in the shortest time the hardware allows.

10.6.3 IT load as a control variable

Every actuator so far removes heat. A data center can also reduce the heat at its source, because the IT equipment itself is controllable. The control input then includes an IT component u_{IT}: power-capping GPUs or processors, which limits their power at some cost in speed; migrating workload to other rooms or sites; deferring batch jobs that have no deadline in the next hours; and shutting down noncritical servers. Since nearly all IT power becomes heat in the room, each megawatt of IT power removed is a megawatt less that the cooling must carry, and the reduction acts within seconds of the command, faster than any chiller can start.

The full decision then chooses the cooling inputs u_{cool} and the IT inputs u_{IT} together,

$$\begin{aligned} &\min_{u_{\text{cool}},\, u_{\text{IT}}} \; \left[\, \text{energy cost}(u_{\text{cool}}, u_{\text{IT}}) + \text{performance penalty}(u_{\text{IT}}) \,\right] \\ &\text{subject to} \quad T_i \le T_{\max} \text{ for all } i, \end{aligned} \tag{10.31}$$

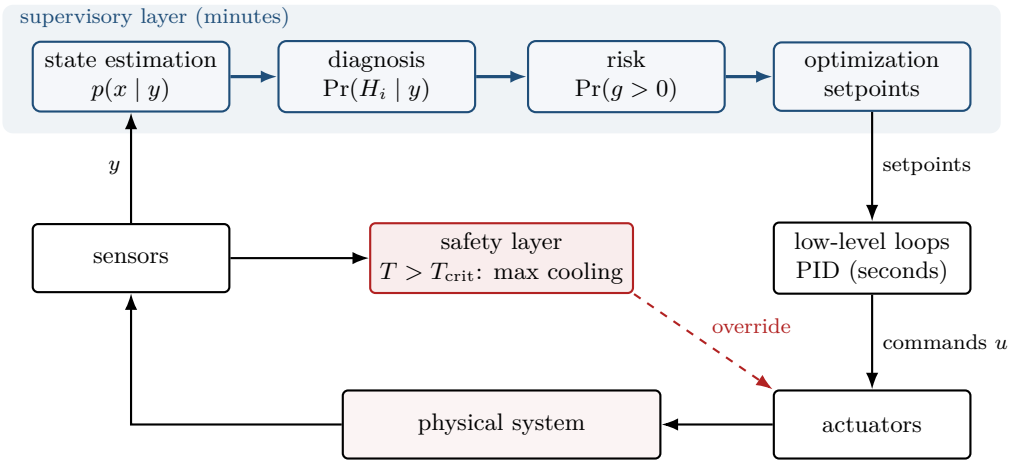


Figure 10.9: Layers of a data center control system. The supervisory layer (top row) estimates, diagnoses, assesses risk and chooses setpoints; the low-level loops hold the setpoints; the safety layer reads the sensors directly and overrides the actuators when a temperature passes T_{crit}.

where the performance penalty prices the slower jobs, the deferred work or the migrated load. Joint optimization lets the facility trade a small performance cost for a large reduction in risk when the cooling is short, and it makes the IT load, normally the largest disturbance of equation (10.4), partly a control input. That changes the boundary between disturbance and control, and it requires an agreement with the owners of the workload on which parts of it are flexible.

10.6.4 The emergency hierarchy

When a chiller fails, the controls respond in an order that uses the cheapest and least disruptive actions first and the most disruptive last. Table 10.4 lists the six steps; each row gives the action, the part of the system it acts on, and its cost. The first two use cooling capacity the plant already has; the third spends thermal margin; the last three act on the IT load, in order of increasing disruption to the service. The order follows from equation (10.31): each step is taken only when the preceding ones cannot hold the constraints, because each carries a larger performance penalty.

Table 10.4: The emergency hierarchy after a chiller failure, in order of use.

Step	Action	Acts on	Cost
1	raise the loading of the remaining chillers	cooling plant	energy; less margin against a second failure
2	raise airflow and pump flow	air and water distribution	fan and pump energy, cubic in speed
3	raise the allowable temperature temporarily	setpoints, limits	less thermal margin; equipment within its allowable but not recommended range
4	power-cap selected servers	IT load	slower computation
5	shift flexible workload elsewhere	IT load	migration effort, latency
6	shed noncritical IT load	IT load	lost service for that load

10.6.5 The compact form

The chapter's controllers are all special cases of one statement. The system obeys equation (10.1), $\dot{x} = f(x, u, d, \theta)$, and the control is

$$\begin{aligned} u^* &= \arg\min_u \; J(x, u) \\ \text{subject to} \quad & g(x, u) \le 0 \quad \text{and possibly} \quad \Pr\big(g(x, u) > 0\big) < \epsilon, \end{aligned} \tag{10.32}$$

where J is the objective accumulated over the horizon (energy, deviation, wear and performance), $g \le 0$ collects the constraints written so that a positive value is a violation, and ϵ is the accepted risk level. A PI loop is the case with one state, no explicit objective and the setpoint as an implicit constraint; model predictive control is the deterministic version over a horizon; the chance constraint and the expectation in equation (10.29) make it stochastic. Equation (10.32) is optimal control under uncertainty, and the rest of the chapter is a list of ways to approximate it with less information or less computation.

10.6.6 The supervisory layer

Equation (10.32) is not solved by replacing the PID loops. The low-level loops stay where they are and do what they do well: hold a fan speed, a valve position, a supply temperature. A supervisory layer sits above them, as in Figure 10.9, and runs the chain estimation, diagnosis, risk and recommended action. Its output is a ranked set of actions, each with the risk it leaves, rather than a command to a single actuator.

Table 10.5 gives an example of its output during an outage of one chiller. The first row is the present situation: if chiller 3 stays unavailable, the probability of violating a rack inlet limit within the next 20 minutes is 12 %. The other rows are two candidate actions and the probability each leaves. Raising the chilled-water pump speed reduces the probability to 8 %; shifting 1 MW of flexible IT load reduces it to 1.5 %. Each probability is a statement about the uncertain future state, computed from the estimate, the forecast and the model; none is measured. The table states the trade-off in a form an operator can act on: the pump action is cheap and removes a third of the risk, the IT shift is more disruptive and removes nearly all of it. Neither action alone brings the risk below the 0.01 level of equation (10.26); if that is the facility's standard, the supervisor's report says so and points to a combination of actions or to a decision to accept 1.5 % for the duration of the outage.

Table 10.5: An example of supervisory output during a chiller outage: candidate actions and the probability of violating a rack inlet limit in the next 20 minutes.

Situation or action	Probability of violation	Risk removed
chiller 3 stays unavailable, no action	12 %	–
raise chilled-water pump speed	8 %	4 points, a third
shift 1 MW of flexible IT load	1.5 %	10.5 points, about 88 %

10.6.7 Control as decision under uncertainty

Control is the choice of u so that the system behaves as wanted. In its easy form the state is measured, the model is known, the disturbance is small, and a PI loop with sensible gains is the whole answer; most of the loops in a cooling plant live in that form and should stay there.

The hard form is choosing u when the state, the parameters, the loads and the future disturbances are all uncertain. Then the controller needs state estimation to know where the

system is, forecasting to know where the disturbances are going, optimization to choose among the actions, and risk management to say how much the choice can be trusted. The supervisory layer of Section 10.6.6 is that combination in operation, and its value in a data center lies in the margins it can safely give up and the failures it can see coming.

10.7 What to retain

Definition 10.1 (Control under uncertainty)**.** Control is the choice of the inputs u of $\dot{x} = f(x, u, d, \theta)$ so that the system behaves as wanted; in a data center that choice is made by low-level feedback loops holding setpoints and, above them, by a supervisory layer that combines state estimation, forecasting, optimization and an explicit risk level, because the state, the parameters, the loads and the future disturbances are uncertain.

Table 10.6 lists the chapter's objects and the subsection that defines each.

Table 10.6: The objects of the control chapter and where each is defined.

Object	Expression	Role	Defined in
Dynamic system	$\dot{x} = f(x, u, d, \theta)$	state, control, disturbance, parameters	10.1.1
Actuators	Table 10.1	how u reaches the plant	10.1.2
Error	$e = T - T_{\text{set}}$ (this chapter)	input of every feedback law	10.2.1
On/off with hysteresis	equation (10.7)	band control, limits switching	10.2.2
Proportional control	$u = K_p e$; $e_{\text{ss}} = \Delta/(1 + bK_p)$	smooth action, steady error	10.2.3
PI and PID	equations (10.13), (10.15)	remove offset; anticipate trend	10.2.4, 10.2.5
Cascade	$T_{\text{rack}} \to \dot{V}_{\text{set}} \to N_{\text{fan}}$	fast inner, slow outer loop	10.3.1
Chiller staging	equation (10.18)	discrete capacity decision	10.3.4
Model predictive control	equation (10.20)	receding-horizon optimization	10.4.2
Cost and constraints	equations (10.21), (10.22)	energy, deviation, wear; safety	10.4.3
Chance constraint	$\Pr(T > T_{\max}) < \epsilon$	risk held at a level	10.5.2
Robust, stochastic control	equations (10.28), (10.29)	worst case; expectation plus risk	10.5.4
Diagnosis-aware action	$\Pr(H_i \mid y)$	action and measurement by cause	10.6.1
Joint IT and cooling	equation (10.31)	IT load as a control input	10.6.3
Compact form	equation (10.32)	optimal control under uncertainty	10.6.5

10.8 Exercises

Exercise 10.1. A zone follows equation (10.10) with $\tau = 8\,\text{min}$, $b = 0.25\,\text{K}$ per percentage point of fan speed, and a load step that would raise it $\Delta = 4\,\text{K}$ above its 24°C setpoint without

control. (a) With proportional control at $K_p = 12$ points per kelvin, compute the loop gain, the steady error, the temperature at which the zone settles and the closed-loop time constant $\tau/(1 + bK_p)$. (b) What gain would limit the steady error to 0.5 K? (c) With an integral term added, how many points of extra fan speed does the controller hold at steady state, and what is the steady error?

Exercise 10.2. An on/off controller cools a zone that warms toward 31°C with the cooling off and cools toward 19°C with it on, both with $\tau = 12$ min. (a) With thresholds $T_{\text{low}} = 23$°C and $T_{\text{high}} = 25$°C, use equation (10.8) to compute the length of the off and on intervals and the number of starts per hour. (b) Repeat with the band narrowed to 23.9–24.1°C. (c) A fan running at 100 % speed draws 15 kW. By the cube law, what does it draw at 80 % and at 70 % speed?

Exercise 10.3. The chance constraint is $\Pr(T > 30\text{°C}) < 0.01$. (a) Rack A has a predicted inlet temperature $\mathcal{N}(26, 1.5^2)$ and rack B $\mathcal{N}(25, 3^2)$, in °C. Compute each rack's probability of exceeding 30°C and say which satisfies the constraint. (b) By how much must the mean of rack B be lowered, at the same standard deviation, to satisfy it? (c) A robust controller models the inlet as $T = T_{\text{supply}} + R_{\text{th}}\dot{Q}_{\text{rack}}$ with $\dot{Q}_{\text{rack}} = 12$ kW and $R_{\text{th}} \in [0.5, 0.9]$ K/kW, and a limit of 32°C. Compute the highest supply temperature it allows, and compare it with the value at the central resistance 0.7 K/kW.

Exercise 10.4. For the rack of Table 10.3, with prior probabilities 0.65, 0.25 and 0.10 for H_1, H_2 and H_3, a second sensor reads high with probability 0.9 under H_1 or H_3 and 0.1 under H_2. (a) Compute the probability that the second sensor reads high. (b) Compute the posterior probabilities of the three hypotheses after a high reading and after a normal reading. (c) Under which reading should the facility raise airflow, and under which should it send a technician to the sensor? Explain why taking the reading before acting is worth a short delay.

10.9 Connection to the TERRA framework

The controls of this chapter act on the same graph that the earlier chapters measured, and the supervisory layer of Section 10.6.6 is the role TERRA plays above the facility's own control loops.

- **The dynamic system is the graph.** The state x of equation (10.2) is the set of node states of three coupled conservation networks: rack and aisle temperatures on thermal-energy nodes, plenum pressure and loop flow on air and water mass nodes, chilled-water temperatures on the loop nodes. Forward simulation of that graph in time is $\dot{x} = f(x, u, d, \theta)$, the prediction step of the model predictive control of Figure 10.7.
- **Actuators are component states with bounds.** Fan speed, valve opening, pump speed, chiller loading and the number of chillers running are states of components on the edges that carry air, water and heat; their ranges are the component bounds of equation (10.22), and the cube law and the COP curve of Table 10.2 are component relations.
- **Estimation before control is posterior inference.** The distribution $p(x \mid y)$ of equation (10.24) is the posterior the framework computes over unmeasured states and parameters: rack airflow, recirculation fraction and coil conductance are inferred, inlet temperatures measured, coil duties calculated. Its standard deviations set the margins of equation (10.27).

- **Risk is a probability of a constraint violation.** The chance constraint and the 12, 8 and 1.5 % of Table 10.5 are probabilities that required cooling exceeds available cooling at some rack within a horizon, computed by propagating the posterior forward through the graph under each candidate action.
- **Diagnosis-aware action uses the ranked posterior.** $\Pr(H_i \mid y)$ of Table 10.3 is the ranked posterior over hypotheses for one symptom, and the second temperature reading is the next measurement chosen for the information it adds.

In summary, the low-level loops of a data center hold setpoints by feedback and should be left to do so, while the choice of setpoints, staging and IT actions is optimal control under uncertainty, which needs an estimate of the state, a forecast of the disturbances and a probability of violating each limit; the graph model supplies the prediction, the posterior supplies the uncertainty, and the supervisory layer turns both into ranked actions. The next chapter turns from how the facility acts to how its efficiency is measured, through power usage effectiveness.

Chapter 11

PUE and energy efficiency: what the ratio measures and what it does not

11.1 Power usage effectiveness

The preceding chapters followed electrical energy into the IT equipment, out of it as heat, through the air and water loops, and into the environment, and then asked what the facility measures of that chain and how it acts on it. This chapter asks how the efficiency of the whole chain is reported. The question is

how efficiently does the facility convert the electrical power it draws into power delivered to the IT equipment, and what does the standard number for that conversion reveal and conceal?

The standard number is the power usage effectiveness, a single ratio of two electrical powers. It is simple, widely reported and useful, and it is easy to read more into it than it contains. This section defines it, works the basic examples and explains why it became the industry's headline metric; the sections that follow say what it leaves out, why it moves, what lies underneath it, how it behaves in time, and how it enters capacity and operating decisions.

Jargon: PUE, overhead, non-IT power, economizer hours, stranded capacity, annual and instantaneous PUE

Power usage effectiveness (PUE) is the ratio of the total electrical power drawn by a facility to the electrical power delivered to its IT equipment, $\mathrm{PUE} = P_{\mathrm{facility}}/P_{\mathrm{IT}}$. *Non-IT power* is everything in P_{facility} that does not reach the IT equipment: cooling, electrical conversion and distribution losses, lighting, offices and other building loads. *Overhead* is the same quantity regarded as the price paid for delivering IT power, often quoted as a fraction of P_{IT}; an overhead of 0.2 means 0.2 MW of non-IT power per megawatt of IT power. An *economizer* is a cooling mode that uses cold outdoor air or cold water from a cooling tower directly, with the chiller compressors off or unloaded; it is also called *free cooling*. *Economizer hours* are the hours of the year in which the weather allows that mode. *Stranded capacity* is infrastructure capacity, electrical or cooling, that is installed but cannot be used because another subsystem limits the IT load first. *Instantaneous PUE* is the ratio of the two powers at one moment; *annual PUE* is the ratio of the two energies over a year, which is the figure normally published.

Key idea

PUE measures one thing, the facility power spent per unit of IT power. It does not measure computing efficiency, energy use, water, carbon, thermal safety, reliability or capacity, and it moves with load and weather when nothing is wrong. Read against a physical model of the facility it becomes evidence about the system; read alone it is a ratio.

11.1.1 The ratio and its two forms

Power usage effectiveness is defined as

$$\mathrm{PUE} = \frac{P_{\text{facility}}}{P_{\text{IT}}}, \tag{11.1}$$

where P_{facility} is the total electrical power entering the facility, normally measured at the utility meter, and P_{IT} is the electrical power consumed by the IT equipment, normally measured at the output of the uninterruptible power supplies (UPS) or at the power distribution units (PDU) that feed the racks. The facility power is the IT power plus everything else,

$$P_{\text{facility}} = P_{\text{IT}} + P_{\text{cooling}} + P_{\text{elec. losses}} + P_{\text{lighting}} + P_{\text{other}} = P_{\text{IT}} + P_{\text{non-IT}}, \tag{11.2}$$

and dividing equation (11.2) by P_{IT} gives the second form of the ratio,

$$\mathrm{PUE} = 1 + \frac{P_{\text{non-IT}}}{P_{\text{IT}}}. \tag{11.3}$$

The first form reads PUE as a multiplier, the number of watts the utility delivers for each watt the servers consume. The second reads it as one plus the overhead fraction, and it is the more useful of the two for analysis, because it isolates the part that the facility can change. The lower bound is $\mathrm{PUE} = 1$, reached only if every watt that enters the facility reaches the IT equipment, with no cooling, no conversion losses and no building loads. That bound is not attainable: heat generated by the IT equipment must be removed, and power conversion always loses something, so $\mathrm{PUE} > 1$ in every real facility.

PUE is a *calculated* quantity, the quotient of two *measured* ones, and its value depends on where the two meters sit. If P_{IT} is metered at the UPS output, the losses of the PDUs and of the cables to the racks are counted as IT power; if it is metered at the rack inlets, they are counted as overhead. The same facility at the same moment can therefore report two different PUE values, and a comparison between facilities is meaningful only when both draw the measurement boundary at the same place. That caveat applies to every number in this chapter.

11.1.2 A facility at 1.20

Take an illustrative facility with $P_{\text{IT}} = 10\,\text{MW}$ and $P_{\text{facility}} = 12\,\text{MW}$. Equation (11.1) gives

$$\mathrm{PUE} = \frac{12\,\text{MW}}{10\,\text{MW}} = 1.20,$$

and equation (11.2) gives the non-IT power, $P_{\text{non-IT}} = 12 - 10 = 2\,\text{MW}$. For every megawatt delivered to the IT equipment, the facility spends another 0.2 MW on cooling, conversion losses, pumps, fans and the rest of the supporting infrastructure: an overhead of 20 % of the IT power.

The same numbers read as energy over a year at constant load give $10\,\text{MW} \times 8760\,\text{h} = 87\,600\,\text{MWh}$ of IT energy and 17 520 MWh of overhead energy. At an illustrative price of \$80 per megawatt-hour, the overhead costs about \$1.4 million a year. Each 0.01 of PUE at this scale is 0.1 MW of continuous power, 876 MWh a year, or about \$70 000 a year at the same price. The second decimal of a PUE is therefore not a rounding detail; it is an operating cost.

11.1.3 Two facilities with the same IT load

Consider two illustrative facilities that host the same IT load, $P_{\mathrm{IT}} = 10\,\mathrm{MW}$. Facility A draws $P_{\mathrm{facility}} = 15\,\mathrm{MW}$ and facility B draws $P_{\mathrm{facility}} = 12\,\mathrm{MW}$:

$$\mathrm{PUE}_A = \frac{15}{10} = 1.50, \qquad \mathrm{PUE}_B = \frac{12}{10} = 1.20.$$

Facility B needs 3 MW less supporting power to deliver the same IT power, 5 MW of overhead against 2 MW; over a year at constant load that is $3 \times 8760 = 26\,280\,\mathrm{MWh}$, and at \$80/ MWh about \$2.1 million.

B is probably the more efficient infrastructure, and for a first comparison the conclusion is reasonable. It is not the whole story. The comparison holds the IT load fixed and says nothing about how much computing either facility delivers with its 10 MW, about the climate each sits in, about the water each consumes, about the carbon intensity of the electricity each buys, about how close either runs to a thermal limit, or about whether the two meter P_{IT} at the same place. The next section takes these omissions one at a time.

11.1.4 Why the metric became popular

PUE is easy to read. PUE = 2.0 means one megawatt of overhead for every megawatt of IT power: half the electricity bill buys no computing. PUE = 1.2 means 0.2 MW of overhead per megawatt of IT. The number is dimensionless, needs only two meters, and maps directly onto the cost and the capacity of the supporting infrastructure. It gives operators, owners and customers a compact, shared measure of infrastructure energy overhead, and it made the efficiency of cooling and power distribution visible at a time when facilities with PUE near 2 were common and nobody was asked to report the figure.

The same properties limit it. A single number from two meters cannot carry information about what the IT equipment does with its power, where in the facility the overhead is spent, or how close the facility runs to its limits. The rest of the chapter treats PUE as what it is: a compact summary of one balance, equation (11.2), useful as long as it is not mistaken for a description of the system.

11.2 What the ratio does not measure

PUE describes the conversion of facility power into IT power and nothing else. This section lists what falls outside it: the efficiency of the computing itself, the absolute amount of energy used, the direction of causation behind a change in a ratio, water, carbon, and the useful work delivered per unit of energy.

11.2.1 Facility efficiency and computing efficiency

Suppose facilities A and B both have $P_{\mathrm{IT}} = 10\,\mathrm{MW}$ and the same PUE, say 1.30, so each draws 13 MW. Over a day, facility A processes 1 000 000 jobs and facility B processes 500 000 jobs of the same kind; the time base is an illustrative assumption. The daily facility energy of each is $13\,\mathrm{MW} \times 24\,\mathrm{h} = 312\,\mathrm{MWh}$, so

$$\text{A: } \frac{312\,000\,\mathrm{kWh}}{1\,000\,000} = 0.312\,\mathrm{kWh\ per\ job}, \qquad \text{B: } \frac{312\,000\,\mathrm{kWh}}{500\,000} = 0.624\,\mathrm{kWh\ per\ job}.$$

Facility B spends twice the energy per unit of work, and the two have identical PUE. The difference lies entirely inside P_{IT}: older or less efficient processors, poor consolidation of workloads, many idle servers drawing their idle power, or software that does more computation for the same result.

PUE cannot see any of this, because it treats the IT power as the useful output. A watt delivered to an idle server and a watt delivered to a fully loaded, efficient accelerator count the same in the denominator. PUE is a measure of the infrastructure that surrounds the IT equipment, not of the computational efficiency of the IT equipment.

11.2.2 Higher IT load and a lower ratio

Because IT power sits in the denominator, adding IT power while the overhead stays nearly constant lowers PUE, even when the added IT power does no useful work. Take an illustrative facility at $P_{\mathrm{IT}} = 5\,\mathrm{MW}$ with 2 MW of overhead:

$$\mathrm{PUE} = \frac{5+2}{5} = \frac{7}{5} = 1.40.$$

The IT load then doubles to $P_{\mathrm{IT}} = 10\,\mathrm{MW}$, for example because inefficient servers are added or existing servers are kept busy with low-value work, and the overhead rises only to 3 MW, because the fixed part of the cooling and distribution system does not grow with the load:

$$\mathrm{PUE} = \frac{10+3}{10} = \frac{13}{10} = 1.30.$$

PUE improved by 0.10 while total facility power rose from 7 MW to 13 MW, an increase of 86 %.

Figure 11.1 draws the two cases as stacked bars. The horizontal axis separates the two operating points; the vertical axis is power in megawatts. The gray part of each bar is the IT power and the red part is the overhead, so the height of each bar is the facility power, written above it with the PUE. The eye reads the height: the right bar is almost twice as tall as the left. The ratio reads only the proportion of red to gray, which is smaller on the right. A lower PUE does not mean lower energy consumption, and a facility that reports only PUE can report an improvement while its electricity bill nearly doubles.

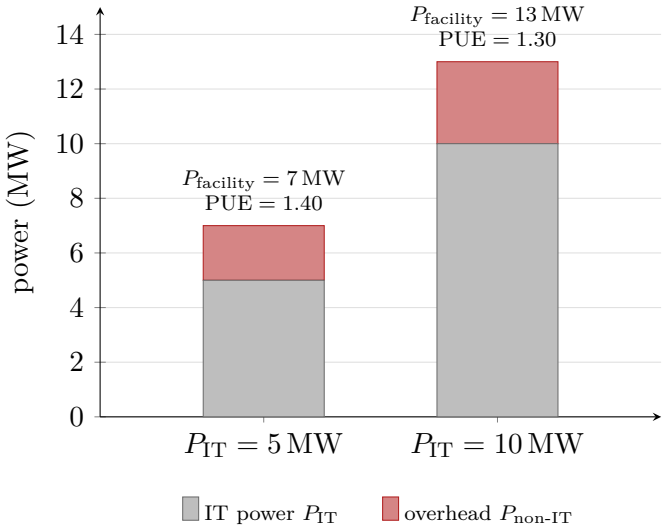


Figure 11.1: An improved PUE with higher energy use, illustrative. Doubling the IT load from 5 to 10 MW while the overhead rises from 2 to 3 MW lowers PUE from 1.40 to 1.30 and raises facility power from 7 to 13 MW.

11.2.3 Reading a ratio

A ratio A/B changes when A falls, when B rises, or when both move at different rates. For PUE in the form of equation (11.3), with $O = P_{\text{non-IT}}$, a small change is

$$\Delta\text{PUE} \approx \frac{\Delta O - (\text{PUE} - 1)\,\Delta P_{\text{IT}}}{P_{\text{IT}}}. \tag{11.4}$$

PUE falls whenever the overhead grows by less than $(\text{PUE} - 1)$ times the growth in IT power. In the example of Section 11.2.2, $\text{PUE} - 1 = 0.4$ and the IT power grew by 5 MW, so any overhead increase below $0.4 \times 5 = 2$ MW lowers PUE; the actual increase was 1 MW. The same reduction of PUE could have come from a genuine cut in overhead at constant IT load, from added IT load at constant overhead, or from any mixture, and the ratio alone does not say which.

The working rule follows: when PUE changes, ask why it changed before concluding that the facility improved or degraded. In practice that means reporting the numerator and the denominator beside the ratio, P_{facility} and P_{IT} in megawatts, so that a change can be attributed to one of them, and, as Section 11.4 develops, reporting the overhead by component.

11.2.4 Water

Evaporative cooling, in cooling towers or in direct evaporative coolers, removes heat by evaporating water. It needs less fan and compressor power than dry cooling at the same duty, so it lowers PUE, and it consumes water that a dry system does not. A facility can therefore achieve a low PUE by spending water, while another facility in a water-stressed region chooses dry coolers or air-cooled chillers, spends more electricity and uses little water.

An illustrative comparison at $P_{\text{IT}} = 10$ MW, with $87\,600$ MWh $= 8.76 \times 10^7$ kWh of IT energy a year: an evaporative facility at PUE = 1.15 that consumes 1.8 liters of water per kilowatt-hour of IT energy uses $1.8 \times 8.76 \times 10^7 \approx 1.58 \times 10^8$ liters, about $158\,000\ \text{m}^3$, a year; a dry-cooled facility at PUE = 1.35 and 0.2 L/kWh uses about $17\,500\ \text{m}^3$ and spends $0.20 \times 87\,600 = 17\,520$ MWh more electricity. Neither is better in general: the trade between electricity and water depends on the local price and scarcity of each. PUE sees only the electricity side, and a separate metric, water per unit of IT energy (often called water usage effectiveness, in liters per kilowatt-hour) or water per unit of workload, is needed for the other.

11.2.5 Carbon

Two facilities with the same PUE can have very different emissions. Take two illustrative facilities at $P_{\text{IT}} = 10$ MW and PUE = 1.20, each drawing 12 MW, 105 120 MWh a year. One buys electricity with an average carbon intensity of 0.05 kg of carbon dioxide equivalent (CO_2e) per kilowatt-hour, from a grid dominated by hydroelectric, nuclear or wind generation; the other buys at 0.70 kg CO_2e/kWh from a grid dominated by coal:

$$1.0512 \times 10^8\ \text{kWh} \times 0.05 \approx 5\,260\ \text{t}\,CO_2e, \qquad 1.0512 \times 10^8\ \text{kWh} \times 0.70 \approx 73\,600\ \text{t}\,CO_2e.$$

The emissions differ by a factor of fourteen at identical PUE. The carbon per unit of IT energy is the product of the two factors, PUE $\times$ (CO_2e per kWh of facility energy), and PUE is only the first. Because grid carbon intensity varies by hour, a facility that shifts flexible load toward low-carbon hours lowers its emissions without changing its PUE at all. PUE is not a measure of carbon efficiency.

11.2.6 Workload efficiency

The quantity a data center exists to produce is useful computing work: trained-model tokens, inference requests, transactions, simulations. The efficiency that matters to its owner is useful

work per unit of facility energy, and it factors into two terms:

$$\frac{\text{useful work}}{E_{\text{facility}}} = \frac{\text{useful work}}{E_{\text{IT}}} \times \frac{E_{\text{IT}}}{E_{\text{facility}}} = \frac{\text{useful work}/E_{\text{IT}}}{\text{PUE}_{\text{annual}}}, \tag{11.5}$$

where E denotes energy over the reporting period. The first factor is the efficiency of the IT equipment and its software, for example tokens of a training run per kilowatt-hour of IT energy or transactions per kilowatt-hour; the second is the infrastructure efficiency, the inverse of PUE. With illustrative numbers, a cluster that produces 2000 tokens per kilowatt-hour of IT energy in a facility at PUE = 1.25 delivers $2000/1.25 = 1600$ tokens per kilowatt-hour of facility energy. Moving the same cluster to a facility at PUE = 1.10 gives 1818, a gain of 14 %; replacing it with hardware that produces 3000 tokens per kilowatt-hour of IT energy in the original facility gives 2400, a gain of 50 %. PUE covers only the facility factor of equation (11.5), and often not the larger one.

11.3 Why PUE varies

PUE is not a constant of a facility. It changes with the IT load, with the weather, with the cooling mode and with the efficiency of the power path, all with no equipment fault. This section derives each dependence and works its example.

11.3.1 Load dependence and the fixed term

Part of the cooling power does not scale with the IT load. Pumps circulate water through the plant at a minimum flow, air handlers keep a minimum airflow, chillers have a minimum stable loading, controls and lighting run regardless. A simple model splits the cooling power into a fixed part P_0 and a part proportional to the IT power,

$$P_{\text{cool}} = P_0 + k\,P_{\text{IT}}, \tag{11.6}$$

with k the incremental cooling power per unit of IT power. If, for simplicity, all non-IT power is lumped into P_{cool}, then $P_{\text{facility}} = P_{\text{IT}} + P_0 + kP_{\text{IT}}$ and

$$\text{PUE} = \frac{P_{\text{IT}} + P_0 + kP_{\text{IT}}}{P_{\text{IT}}} = 1 + k + \frac{P_0}{P_{\text{IT}}}. \tag{11.7}$$

The ratio has a floor, $1 + k$, approached at high load, and a term P_0/P_{IT} that grows without bound as the IT load falls. At low utilization the fixed term dominates and PUE is poor, even though no component is degraded and the cooling system is doing exactly what it was designed to do. A newly opened facility that is a fifth full, or a facility whose load has moved elsewhere, reports a poor PUE for this reason alone.

11.3.2 The same cooling system at two loads

Take the illustrative cooling model $P_{\text{cool}} = 1 + 0.1\,P_{\text{IT}}$, with powers in megawatts, so $P_0 = 1\,\text{MW}$ and $k = 0.1$. At $P_{\text{IT}} = 2\,\text{MW}$,

$$P_{\text{cool}} = 1 + 0.1 \times 2 = 1.2\,\text{MW}, \qquad P_{\text{facility}} = 2 + 1.2 = 3.2\,\text{MW}, \qquad \text{PUE} = \frac{3.2}{2} = 1.6.$$

At $P_{\text{IT}} = 10\,\text{MW}$,

$$P_{\text{cool}} = 1 + 0.1 \times 10 = 2\,\text{MW}, \qquad P_{\text{facility}} = 10 + 2 = 12\,\text{MW}, \qquad \text{PUE} = \frac{12}{10} = 1.2.$$

The same cooling system, with the same equipment and the same control settings, reports 1.6 at one load and 1.2 at another. A comparison of PUE between the two operating points, or between two facilities at different utilizations, compares loads as much as it compares equipment.

Figure 11.2 draws equation (11.7) for this cooling model. The horizontal axis is the IT load in megawatts and the vertical axis is PUE. The solid curve is the PUE; the dashed horizontal line is its floor, $1 + k = 1.10$; the shaded band between them is the fixed term P_0/P_{IT}, the part of the PUE that is due only to running below full load. The two marked points are the examples above. The curve falls steeply below about 4 MW and flattens above about 10 MW: most of the load dependence lives at low utilization, where a small change in load moves PUE by tenths.

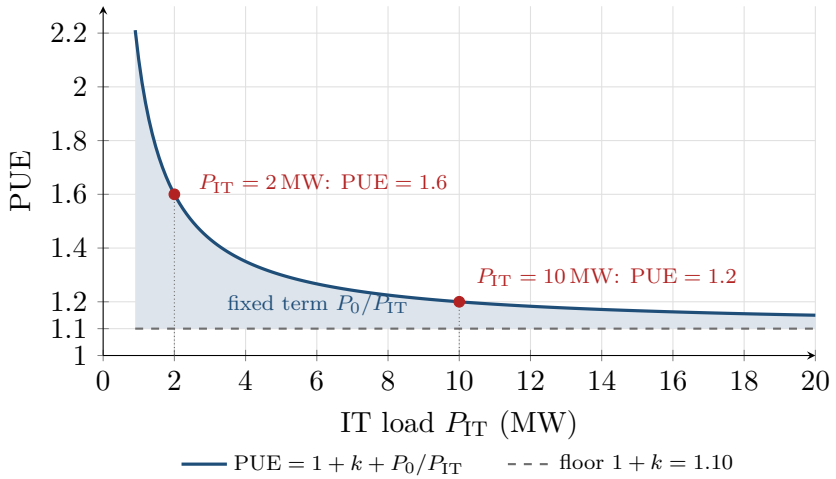


Figure 11.2: PUE against IT load for the illustrative cooling model $P_{\mathrm{cool}} = 1 + 0.1\,P_{\mathrm{IT}}$ (MW), with all non-IT power lumped into cooling, so PUE $= 1.1 + 1/P_{\mathrm{IT}}$ (equation (11.7), $P_0 = 1$ MW, $k = 0.1$). The shaded band is the fixed term. The same system reports 1.6 at 2 MW and 1.2 at 10 MW.

Principle 11.1. *PUE depends on the IT load through the fixed part of the overhead,* $\mathrm{PUE} = 1 + k + P_0/P_{\mathrm{IT}}$*: a facility at low utilization reports a poor PUE with nothing broken.*

11.3.3 Weather

Cooling rejects heat to the outdoors, so its power depends on outdoor conditions. A chiller's compressor works against the temperature difference between the chilled water and the condenser; the condenser is cooled by a cooling tower whose leaving-water temperature follows the outdoor wet-bulb temperature T_{wb}, or by outdoor air at the dry-bulb temperature T_{amb}. Hotter or more humid weather raises the condenser temperature, lowers the chiller's coefficient of performance and raises its compressor power for the same heat removed. In general,

$$P_{\mathrm{cool}} = f(P_{\mathrm{IT}}, T_{\mathrm{amb}}, T_{\mathrm{wb}}, \mathrm{RH}), \qquad \mathrm{PUE} = f(P_{\mathrm{IT}}, \text{weather}, \text{equipment state}), \tag{11.8}$$

where the equipment state includes the number of chillers running, the setpoints and the efficiency of each component.

The consequence is that PUE has a daily and a seasonal cycle. An illustrative facility reports PUE = 1.30 on a hot, humid afternoon and 1.12 the following cool night, at the same IT load. The design did not change overnight and no equipment was repaired; the weather changed. A

comparison between two facilities in different climates, or between this facility's summer and winter, is in part a comparison of weather.

Figure 11.3 draws an illustrative closed form of this dependence at fixed IT load. The horizontal axis is the outdoor temperature in degrees Celsius; the vertical axis is PUE. The three shaded bands are the three cooling modes. Below 12°C the economizer carries the whole load and the compressors are off, so PUE is flat at 1.08. Between 12 and 18°C the economizer carries part of the load and the chillers the rest, and PUE rises steeply, by 0.02 per kelvin, as the compressors take over. Above 18°C the chillers carry the whole load and PUE rises more slowly, by 0.01 per kelvin, as the condenser temperature rises with the weather. The two marked points are the cool night at 14°C and the hot afternoon at 28°C. The figure uses one temperature for simplicity; for a water-side economizer and a cooling tower the governing variable is the wet-bulb temperature, and humidity shifts the curve.

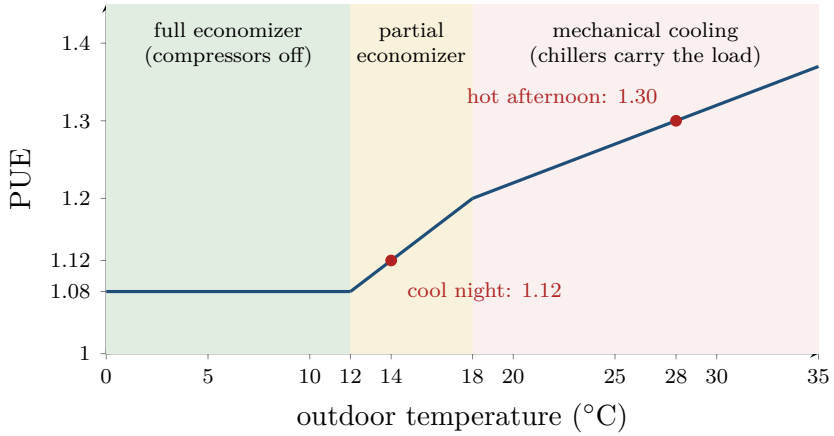


Figure 11.3: PUE against outdoor temperature at fixed IT load, illustrative closed form: $\mathrm{PUE} = 1.08$ for $T \le 12°\mathrm{C}$, $1.08 + 0.02\,(T - 12)$ for $12 < T \le 18°\mathrm{C}$, and $1.20 + 0.01\,(T - 18)$ above 18°C. At $P_{\mathrm{IT}} = 10\,\mathrm{MW}$ the overhead is 0.8 MW with the full economizer and 2 MW at the start of mechanical cooling.

11.3.4 Economizers

An economizer cools with the outdoors directly whenever the outdoors is cold enough. An air-side economizer brings filtered outdoor air into the room or through an air-to-air heat exchanger; a water-side economizer cools the chilled-water loop through a plate heat exchanger from the cooling-tower water, bypassing the chiller. Either way the compressors, the largest single consumers of cooling power, are unloaded or switched off, and only fans and pumps remain. In the illustrative facility of Figure 11.3, at $P_{\mathrm{IT}} = 10\,\mathrm{MW}$, the overhead falls from 2 MW in mechanical cooling ($\mathrm{PUE} = 1.20$) to 0.8 MW with the full economizer ($\mathrm{PUE} = 1.08$), a reduction of 60 %.

The annual PUE of a facility therefore depends on how many hours of the year its climate offers each mode, which is why climate is a siting criterion. With the closed form of the figure, constant IT load, and illustrative hour counts, a cool site with 5000 full-economizer hours at 1.08, 2500 partial hours averaging 1.14 and 1260 mechanical hours averaging 1.25 has

$$\mathrm{PUE}_{\mathrm{annual}} = \frac{5000 \times 1.08 + 2500 \times 1.14 + 1260 \times 1.25}{8760} = \frac{9825}{8760} \approx 1.12,$$

while a hot site with 1000, 2000 and 5760 hours in the same modes, the mechanical hours averaging 1.27, has

$$\mathrm{PUE}_{\mathrm{annual}} = \frac{1000 \times 1.08 + 2000 \times 1.14 + 5760 \times 1.27}{8760} = \frac{10\,675}{8760} \approx 1.22.$$

The same design reports 0.10 more PUE at the hot site because of its weather. Raising the temperatures at which the facility operates, warmer supply air and warmer chilled water, moves the boundaries of Figure 11.3 to the right and adds economizer hours; Section 11.6.5 returns to that lever and its cost.

11.3.5 Electrical losses

Not all of the overhead is cooling. Power on its way from the utility to the servers passes through transformers, switchgear, UPS converters, PDUs and cables, and each stage loses a fraction of what passes through it. For an IT load of $P_{\mathrm{IT}} = 10\,\mathrm{MW}$ supplied through a UPS with efficiency $\eta_{\mathrm{UPS}} = 0.97$, the UPS input power is

$$P_{\mathrm{UPS,in}} = \frac{P_{\mathrm{IT}}}{\eta_{\mathrm{UPS}}} = \frac{10}{0.97} = 10.31\,\mathrm{MW},$$

so 0.31 MW is lost in the UPS alone before any cooling is counted, 0.031 of PUE. A transformer with efficiency 0.99 upstream raises the input to $10.31/0.99 = 10.41\,\mathrm{MW}$, another 0.10 MW, and conductors and other power electronics add more. The efficiency of a UPS also falls at low load fraction, so lightly loaded redundant power paths lose a larger share of what they carry.

These losses appear twice in the facility power. They are drawn from the utility as electrical power, and they become heat inside the electrical rooms, which the cooling system must then remove with its own power. PUE captures the combined overhead of the power path and the cooling, without separating them; equation (11.2) holds them as distinct terms, and the next section uses that distinction.

11.4 Decomposition and local problems

A single ratio says how much overhead there is; it does not say where the overhead is spent or what it costs in thermal safety and resilience. This section breaks the overhead into components, shows that a good PUE can hide a dangerous local condition, and replaces the goal of minimizing PUE with an objective that prices risk.

11.4.1 Component-level overhead

The non-IT power of equation (11.2) is itself a sum of components,

$$P_{\text{non-IT}} = P_{\text{fans}} + P_{\text{pumps}} + P_{\text{chillers}} + P_{\text{towers}} + P_{\text{UPS loss}} + P_{\text{transformer loss}} + \cdots, \tag{11.9}$$

and each term belongs to a subsystem with its own physics, its own controls and its own owner. A PUE that has risen by 0.03 demands an answer to the question *which subsystem is responsible*, and only the decomposition can answer it.

The decomposition is more actionable than PUE alone because each component has its own efficiency metric and its own lever. Chillers are judged by kilowatts of electrical power per kilowatt of heat removed, the inverse of the coefficient of performance, and respond to chilled-water and condenser-water temperatures; fans follow the cube of their speed and respond to airflow setpoints and containment; pumps follow the same affinity laws and respond to differential-pressure setpoints; UPS losses depend on the operating mode and the load fraction. Equation (11.9) requires submeters on the main consumers. Where a component is not metered, its power is calculated from a model or inferred from the balance, and the difference between the sum of the submeters and the main meter is itself a check on the meters.

11.4.2 A decomposition of a 13 MW facility

Take an illustrative facility with $P_{\text{IT}} = 10\,\text{MW}$ and $P_{\text{facility}} = 13\,\text{MW}$, so PUE = 1.30 and 3 MW of overhead. Table 11.1 lists the overhead by component, with each component's share of the overhead and its contribution to PUE, which is its power divided by P_{IT}. The six components sum to 3.0 MW, matching the meter difference $13 - 10$, and their contributions sum to 0.30.

Table 11.1: Overhead of the illustrative 13 MW facility by component. The contribution to PUE is the component's power divided by $P_{\text{IT}} = 10\,\text{MW}$.

Component	Power (MW)	Share of overhead	Contribution to PUE
Chillers (compressors)	1.2	40 %	0.12
CRAH fans	0.5	17 %	0.05
Pumps	0.4	13 %	0.04
Cooling towers	0.2	7 %	0.02
UPS and transformer losses	0.5	17 %	0.05
Other (lighting, offices, controls)	0.2	7 %	0.02
Total overhead	3.0	100 %	0.30

Figure 11.4 draws the same numbers. The left bar is the whole facility power, 13 MW, drawn to scale: the gray block is the IT power and the colored cap is the overhead. The right bar magnifies the cap to show its six components, stacked in the order of the table with the chillers at the bottom, with each component's power and share written beside it. The dashed lines connect the cap to its magnification.

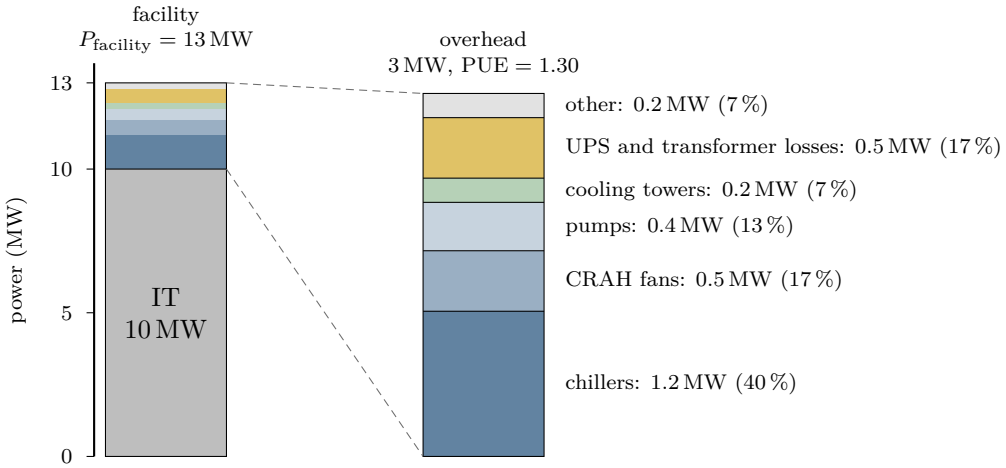


Figure 11.4: Facility power of the illustrative 13 MW facility (Table 11.1). Left: IT power and overhead to scale. Right: the 3 MW overhead magnified four times and split by component. The chillers are the largest single term.

The chillers dominate, at 1.2 MW or 40 % of the overhead, and they are the largest single opportunity. A 20 % reduction in chiller power, from a warmer chilled-water setpoint or more economizer hours, saves 0.24 MW and lowers PUE from 1.30 to 1.276; the same 20 % reduction in fan power saves 0.10 MW, and in tower power 0.04 MW. A fan saving is not small in absolute terms, but the ranking tells an operator where to look first, and PUE alone cannot produce the ranking.

11.4.3 A good PUE with a hot rack

PUE is a facility-wide average and can hide a local problem. Suppose a facility reports an excellent PUE = 1.15, achieved in part by reducing the speed of the air-handler fans to save fan power. In one row, the reduced airflow no longer meets the demand of a high-density rack: hot exhaust recirculates over the top of the rack into its intakes, and its inlet temperature reaches 34°C, several kelvin above the upper end of the recommended inlet range of about 27°C that most operators target. The rack's servers raise their own fan speeds, throttle their processors, or, if the condition persists, shut down.

None of this moves PUE appreciably. One rack of 30 kW in a 10 MW facility is 0.3 % of the IT load, and its extra server-fan power, a few kilowatts, changes PUE in the fourth decimal. An excellent PUE does not imply safe thermal operation; it may even be the cause of the unsafe operation, when the energy saving was bought with airflow.

11.4.4 Efficiency against resilience

Consider two illustrative chiller-staging policies for a facility with a cooling load of $\dot{Q}_{\text{req}} = 10\,\text{MW}$ and chillers of 6 MW capacity each. Under policy A, three chillers run conservatively at about 3.3 MW each, 56 % of capacity; the facility reports PUE = 1.25, because each running chiller brings its own condenser pump, primary pump and tower fan, and keeps spare capacity. Under policy B, two chillers run near their limits at 5 MW each, 83 % of capacity, and the third is off; with fewer auxiliaries running, the facility reports PUE = 1.18.

The difference appears when a chiller trips. Under policy A the two survivors pick up 5 MW each and the available capacity, $\dot{Q}_{\text{avail}} = 12\,\text{MW}$, still exceeds the requirement. Under policy B the survivor offers $\dot{Q}_{\text{avail}} = 6\,\text{MW}$ against $\dot{Q}_{\text{req}} = 10\,\text{MW}$ until the standby chiller has started and loaded, which takes several minutes; during that interval the room heats on its thermal mass alone. Policy B is more efficient and less resilient. PUE rewards it for exactly the property, running with little spare capacity, that makes it fragile. The trade between energy efficiency and resilience margin is a design and operating decision that PUE cannot make, and a later chapter on redundancy and reliability gives it a quantitative form.

11.4.5 Efficiency against thermal margin

The same trade appears at the scale of a single fan. By the fan affinity laws, airflow scales with speed and fan power with its cube,

$$\dot{V} \propto N_{\text{fan}}, \qquad P_{\text{fan}} \propto N_{\text{fan}}^3. \tag{11.10}$$

Reducing the speed by 20 % reduces fan power to $0.8^3 = 0.512$ of its value, a saving of almost half. It also reduces the air mass flow to 0.8 of its value, and the temperature rise across a rack, $\Delta T = \dot{Q}_{\text{rack}}/(\dot{m}_{\text{air}}\, c_{p,\text{air}})$, increases by $1/0.8 = 1.25$: an illustrative rack with a 12 K rise now has a 15 K rise, and with less supply air the risk of recirculation into the inlets grows.

The operating point is therefore not the lowest fan speed but the lowest fan power that keeps every rack inside its limit with an adequate margin,

$$\min_{N_{\text{fan}}} P_{\text{fan}}(N_{\text{fan}}) \quad \text{subject to} \quad T_{\text{in},i}(N_{\text{fan}}) < T_{\max} \text{ for every rack } i, \text{ with adequate risk margin.} \tag{11.11}$$

The phrase *adequate risk margin* carries the difficulty. The rack inlet temperatures depend on airflow paths that are only partly measured, so the constraint is uncertain, and the next subsection states it as a probability.

11.4.6 A probabilistic comparison

Suppose a model of the facility, propagating the uncertainty in loads, weather and the unmeasured airflow paths, assigns each operating policy both a PUE and a probability of a thermal violation, an event in which at least one rack inlet exceeds its limit during the year. Two illustrative policies:

$$\text{A: PUE} = 1.24, \quad \Pr(\text{violation}) = 0.001; \qquad \text{B: PUE} = 1.18, \quad \Pr(\text{violation}) = 0.04.$$

B saves energy and carries forty times the risk. PUE alone cannot choose between them; the choice is a multi-objective decision.

Figure 11.5 places the two policies in an efficiency–risk plane. The horizontal axis is PUE and the vertical axis is the annual probability of a thermal violation, in percent; lower and to the left is better on both counts. The two dots are the policies. The dashed horizontal line is an illustrative risk limit of 1 %, which B violates. The solid line is the set of policies that have the same combined cost as A when a violation is priced at the break-even value computed below; a policy below and to the left of that line is preferred to A at that price.

The break-even price follows from the energy difference. At $P_{\mathrm{IT}} = 10\,\mathrm{MW}$, 0.06 of PUE is 0.6 MW of continuous power, or $0.6 \times 8760 = 5256\,\mathrm{MWh}$ a year; at \$80/ MWh that is about \$420 000 a year in favor of B. B adds $0.04 - 0.001 = 0.039$ to the annual probability of a violation. B is preferred only if the expected cost of that added probability is smaller than the saving, that is, if a thermal violation costs less than

$$\lambda^{*} = \frac{\$420\,000}{0.039} \approx \$10.8 \text{ million}.$$

For a facility whose violation means lost hardware, breached service agreements and customer departure, the true cost may well exceed that figure, and A is the better policy despite its worse PUE.

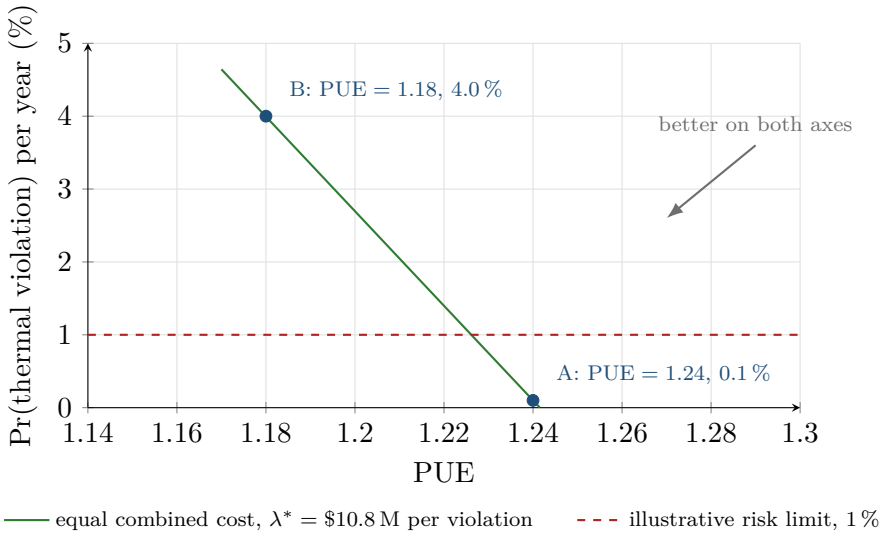


Figure 11.5: The efficiency–risk plane for two illustrative policies. A: PUE = 1.24, annual violation probability 0.001; B: PUE = 1.18, probability 0.04. The solid line joins policies of equal combined cost, annual energy cost plus λ^{*} times the violation probability, when the energy cost is priced at $P_{\mathrm{IT}} = 10\,\mathrm{MW}$ and \$80/ MWh and a violation at $\lambda^{*} \approx \$10.8$ million.

11.4.7 A better objective

The comparison generalizes. The goal of operating a data center is not to minimize PUE but to minimize a total expected cost in which energy is one term and the risks are priced explicitly,

$$\begin{aligned} \min_{u} \;\; & \mathbb{E}\big[C_{\text{energy}}(u)\big] + \lambda_1 R_{\text{thermal}}(u) + \lambda_2 R_{\text{reliability}}(u) + \lambda_3 C_{\text{performance}}(u) \\ \text{subject to} \;\; & \text{operational constraints,} \end{aligned} \tag{11.12}$$

where u collects the operating decisions (setpoints, fan and pump speeds, chiller staging), R_{thermal} is a probability of a thermal violation, $R_{\text{reliability}}$ a probability of losing cooling or power capacity below the requirement after a failure, $C_{\text{performance}}$ a penalty for throttled or delayed workload, and λ_1, λ_2, λ_3 convert each into the units of cost. The constraints hold the hard limits: equipment ratings, temperature and humidity bounds, and any risk limits set by policy, such as the 1 % line of Figure 11.5.

Minimizing PUE is the special case $\lambda_1 = \lambda_2 = \lambda_3 = 0$ at fixed IT load, which assigns no cost to any risk. Every operating decision that trades energy against margin, the fan speed of equation (11.11), the chiller staging of Section 11.4.4, the setpoint change of Section 11.6.5, is an instance of equation (11.12) with nonzero weights.

Principle 11.2. *A good PUE does not imply safe or resilient operation. The decision that PUE alone cannot make is the trade between energy and risk, and it requires both the energy and the probability of a violation to be computed.*

11.5 PUE in time

A facility's PUE is a time series, driven by load and weather. This section distinguishes instantaneous from annual PUE, conditions the series on its drivers, builds the expected PUE and its residual into a detector of degradation, and uses the same model to forecast facility power.

11.5.1 Instantaneous and annual PUE

Two definitions are in use. The instantaneous PUE is the ratio of the two powers at one time,

$$\text{PUE}(t) = \frac{P_{\text{facility}}(t)}{P_{\text{IT}}(t)}, \tag{11.13}$$

usually computed from averages over a short interval such as fifteen minutes. The annual PUE is the ratio of the two energies over a year,

$$\text{PUE}_{\text{annual}} = \frac{E_{\text{facility}}}{E_{\text{IT}}} = \frac{\int P_{\text{facility}}(t)\,\mathrm{d}t}{\int P_{\text{IT}}(t)\,\mathrm{d}t} = \frac{\sum_t P_{\text{IT}}(t)\,\text{PUE}(t)}{\sum_t P_{\text{IT}}(t)}. \tag{11.14}$$

The last form shows that the annual PUE is the average of the instantaneous values weighted by IT power, not their simple average. Hours of low load, which have high PUE by equation (11.7), count for less. With an illustrative year split into two halves, 4380 h at $P_{\text{IT}} = 12\,\text{MW}$ and $\text{PUE} = 1.15$ and 4380 h at $P_{\text{IT}} = 6\,\text{MW}$ and $\text{PUE} = 1.35$, the simple average is 1.25 and the annual PUE is

$$\text{PUE}_{\text{annual}} = \frac{12 \times 1.15 + 6 \times 1.35}{12 + 6} = \frac{13.8 + 8.1}{18} \approx 1.217.$$

The two definitions serve different purposes. The annual value smooths over weather, workload cycles and operational transients, and it is the figure that is published and compared. The instantaneous value shows the operational variation that the annual figure averages away, and it is the one an operator uses to see what the facility is doing now.

11.5.2 Time-series PUE and its drivers

Read as a time series, PUE is a response to a small number of drivers: the IT load $P_{\mathrm{IT}}(t)$, the outdoor dry-bulb temperature $T_{\mathrm{amb}}(t)$, the humidity $\mathrm{RH}(t)$ or the wet-bulb temperature $T_{\mathrm{wb}}(t)$, and the operating mode, such as the number of chillers running $N_{\mathrm{ch}}(t)$ or the fraction of economizer operation. Plotted against time alone, PUE rises and falls with every warm afternoon and every load swing, and a slow deterioration in one component is buried under that variation.

Plotted against its drivers, the same series separates into a part the drivers explain and a part they do not. The explained part is normal operation: PUE at 1.25 on an afternoon at 30°C and 6 MW of load is unremarkable. The unexplained part is where degradation and changed settings appear. The rest of this section makes that separation explicit.

11.5.3 Degradation detection

With enough history, the relation between PUE and its drivers can be learned from a reference period in which the facility is known to have been operating normally, giving a model

$$\mathrm{PUE} = f(P_{\mathrm{IT}}, T_{\mathrm{amb}}, T_{\mathrm{wb}}, \mathrm{mode}). \tag{11.15}$$

If the facility later reports a PUE significantly higher than the model predicts under the same load, weather and mode, something about the facility has changed. Candidate causes include chiller degradation (fouled condenser tubes, refrigerant loss), pump inefficiency (worn impellers, a valve left throttled), fouling of a cooling coil that forces higher fan speeds, a changed control setting such as a lowered chilled-water setpoint, or a bias in one of the two meters.

PUE in this role is no longer a performance score; it is a residual signal. Conditioned on the operating conditions, it measures the departure of the facility from its own normal behavior, and a departure can be detected long before it would be visible in the raw series. The conditioning is what makes the signal: the same rise of 0.03 means nothing on a day that is 5 K warmer and a great deal on a day with the same weather.

11.5.4 The expected PUE and the residual

A rise from 1.18 to 1.25 may be entirely expected if the wet-bulb temperature rose over the same weeks, and an unchanged PUE may hide a fault if the weather cooled. A fixed threshold on PUE cannot tell these cases apart. The quantity to watch is the residual

$$r_{\mathrm{PUE}}(t) = \mathrm{PUE}_{\mathrm{observed}}(t) - \mathrm{PUE}_{\mathrm{expected}}(t), \qquad \mathrm{PUE}_{\mathrm{expected}} = f(P_{\mathrm{IT}}, \mathrm{weather}, \mathrm{mode}), \tag{11.16}$$

where the observed PUE is calculated from the two meters and the expected PUE is calculated from the model of equation (11.15). In normal operation the residual scatters around zero with a spread set by meter noise and model error; a persistent positive residual is the signature of a change in the facility.

Figure 11.6 applies the residual to a synthetic 120-day record of daily means, generated by the script described in the caption. The upper panel shows PUE against day: the dots are the observed values, the dashed curve is the expected PUE, and the red dashed horizontal line is a fixed alarm threshold at 1.22. The expected PUE rises from about 1.19 to about 1.25 over the record because the wet-bulb temperature rises from about 8 to about 20°C; the weekly ripple is the IT load. From day 85 a degradation adds a slowly growing excess, 0.0012 per day, which is known here because the record is synthetic. The lower panel shows the residual of equation (11.16), with the shaded band at ±0.012, three times the standard deviation of the meter and model noise.

The two detectors behave differently. The fixed threshold first fires on day 40 and then fires repeatedly through days 52 to 85, every time the weather is warm or the load is low, while

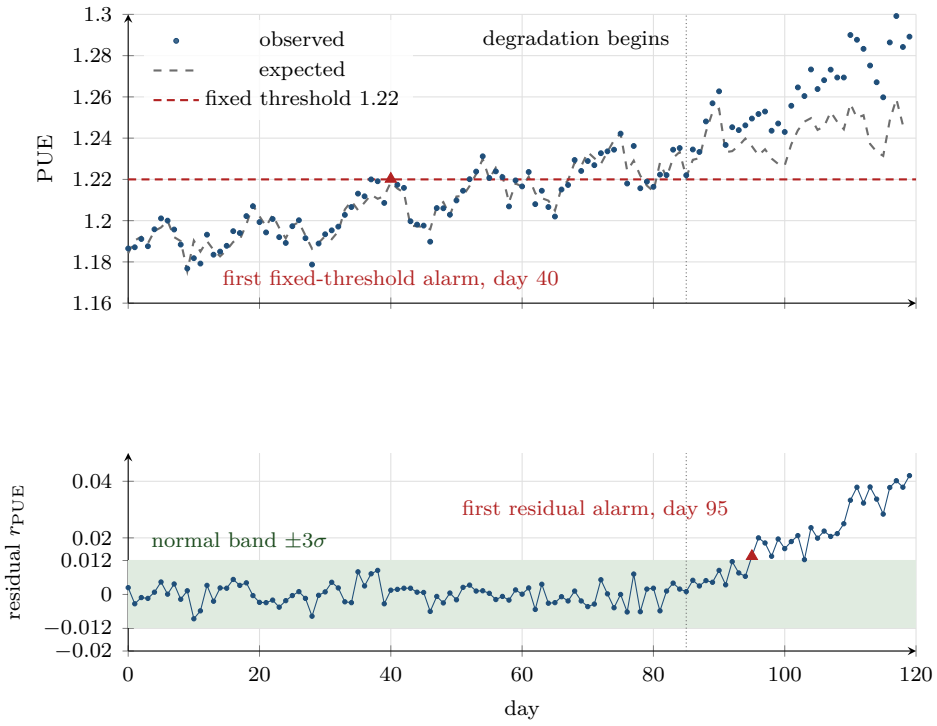


Figure 11.6: Observed against expected PUE over 120 synthetic days, and the residual. Generated by `scripts/ch11_data.py` (seed 11): $P_{\mathrm{IT}} = 20 + 1.2\sin(2\pi d/7)$ MW; T_{wb} rising from 8 to 20°C with noise; $\mathrm{PUE}_{\mathrm{expected}} = 1.07 + 1.4/P_{\mathrm{IT}} + 0.0055\,T_{\mathrm{wb}}$; from day 85 an added degradation $0.0012\,(d - 85)$; meter and model noise of standard deviation 0.004. The fixed threshold fires on day 40 from weather alone; the residual stays inside its band until day 95, ten days after the degradation begins.

the facility is healthy: those are false alarms, and an operator who sees them weekly learns to ignore the threshold. The residual stays inside its band for the first 85 days, because the model explains the weather and the load, and leaves it on day 95, when the degradation has accumulated to $0.0012 \times 10 = 0.012$; by the last ten days its mean is about 0.036, or 0.72 MW of unexplained overhead at $P_{\mathrm{IT}} = 20\,\mathrm{MW}$. The residual is more useful than any fixed threshold because it compares the facility with itself under the same conditions.

11.5.5 Forecasting facility power

The model of equation (11.15) also runs forward. A forecast of the IT load $P_{\mathrm{IT}}(t+h)$ and of the weather $T_{\mathrm{amb}}(t+h)$ and $T_{\mathrm{wb}}(t+h)$ at a horizon h gives a prediction of the cooling power $P_{\mathrm{cool}}(t+h)$, of the facility power and of the PUE. Each prediction has a use: the facility power forecast feeds energy purchasing and the facility's own load forecast to its utility; the peak of that forecast predicts the demand charge and any approach to the interconnection limit; the hourly profile against electricity prices supports cost optimization, such as precooling before an expensive hour; and the cooling forecast supports staging decisions, such as when to bring a chiller online before the afternoon.

As an illustrative example, the forecast for tomorrow at 15:00 is an IT load of $P_{\mathrm{IT}} = 50\,\mathrm{MW}$, a cooling power of $P_{\mathrm{cool}} = 8\,\mathrm{MW}$ in the afternoon heat, and 2 MW of other overhead (losses, lighting, auxiliaries):

$$P_{\mathrm{facility}} = 50 + 8 + 2 = 60\,\mathrm{MW}, \qquad \mathrm{PUE} = \frac{60}{50} = 1.20.$$

The utility sees 60 MW at the meter, not 50. An operator who plans grid capacity, demand charges or a demand-response commitment from the IT load alone underestimates the facility's draw by 10 MW, 20 %, and the error grows on a hotter afternoon, when PUE is higher.

11.6 Capacity, evidence and decisions

PUE also enters planning and diagnosis. This section shows how it limits the IT capacity under a fixed grid connection, why usable capacity is the minimum over subsystems, how an anomalous PUE becomes evidence about specific causes, and what a physical model can answer that the ratio cannot.

11.6.1 Capacity under an interconnection limit

A facility's grid connection is sized for a maximum power, $P_{\mathrm{grid,max}}$. Since $P_{\mathrm{facility}} = \mathrm{PUE} \times P_{\mathrm{IT}}$, the largest IT load that the connection supports at a given PUE is

$$P_{\mathrm{IT,max}} = \frac{P_{\mathrm{grid,max}}}{\mathrm{PUE}}. \tag{11.17}$$

For an illustrative interconnection limit of 100 MW,

$$\mathrm{PUE} = 1.25: \; P_{\mathrm{IT,max}} = \frac{100}{1.25} = 80\,\mathrm{MW}, \qquad \mathrm{PUE} = 1.10: \; P_{\mathrm{IT,max}} = \frac{100}{1.10} \approx 90.9\,\mathrm{MW}.$$

The efficiency gain releases about 10.9 MW of IT capacity through the same connection, without a new substation. Since grid connections are slow and expensive to enlarge, and IT capacity is what a facility sells, this is often the most commercially important consequence of a lower PUE. The relevant PUE is the one at the peak, the hottest hour at full load, not the annual value, because the connection limit applies at every instant.

Figure 11.7 draws equation (11.17). The horizontal axis is PUE and the vertical axis is the maximum IT load in megawatts for the 100 MW connection. The curve is a hyperbola:

each reduction of PUE releases capacity, and the release per 0.01 of PUE is larger at low PUE, about 0.83 MW per 0.01 near 1.10 against 0.64 MW near 1.25. The two marked points are the examples.

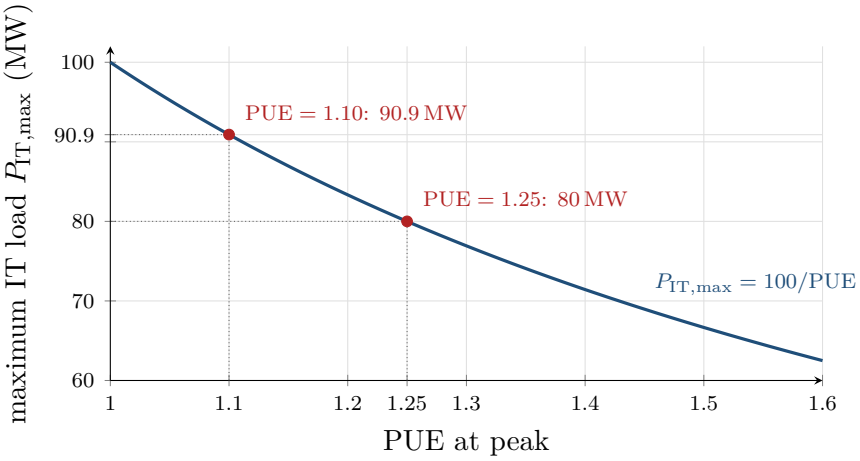


Figure 11.7: Maximum IT load against PUE for an illustrative 100 MW interconnection, $P_{\mathrm{IT,max}} = 100\,\mathrm{MW}/\mathrm{PUE}$ (equation (11.17)). Lowering the peak PUE from 1.25 to 1.10 releases about 10.9 MW of IT capacity.

11.6.2 Usable and stranded capacity

The grid connection is only one limit. Every subsystem supports some maximum IT load: the electrical distribution through its transformers, UPS modules and PDUs; the cooling through its chillers, pumps, air handlers and airflow paths; the floor through its space and rack positions; the network through its ports and uplinks. The IT load the facility can actually host is the smallest of these,

$$C_{\mathrm{usable}} = \min(C_{\mathrm{electrical}},\ C_{\mathrm{cooling}},\ C_{\mathrm{space}},\ C_{\mathrm{network}},\ \ldots), \tag{11.18}$$

with every C expressed in megawatts of IT load. If the electrical system supports 90 MW of IT load and the cooling system only 82 MW, the facility can host 82 MW, whatever its PUE. Usable capacity is a property of the whole system, set by its tightest subsystem, and it moves when any subsystem changes: a failed chiller, a blocked airflow path or a UPS module out for maintenance can lower it.

The capacity that the other subsystems offer above the minimum is stranded: installed, paid for and unusable. With an electrical capacity of 100 MW and a cooling capacity of 80 MW, only 80 MW is usable and 20 MW of electrical capacity is stranded until the cooling is expanded. The same happens at smaller scale with the cooling itself: a facility can have enough total cooling capacity for its load while poor airflow, bypass and recirculation prevent that capacity from reaching particular racks, so that racks in one zone cannot be filled even though the plant has spare capacity.

Figure 11.8 draws equation (11.18) for two subsystems. The horizontal axis is the cooling capacity and the vertical axis is capacity, both in megawatts of IT load, with the electrical capacity fixed at 100 MW (horizontal dashed line). The diagonal dashed line is the cooling capacity itself. The thick curve is the usable capacity, the lower of the two. The shaded triangle on the left is stranded electrical capacity, where cooling is the limit; the shaded triangle on the right is stranded cooling capacity, where electrical is the limit. The marked point is the example: at $C_{\mathrm{cooling}} = 80$ MW the usable capacity is 80 MW and the bracket marks the 20 MW

stranded above it. Usable capacity is maximal, and nothing is stranded, only where the two limits meet.

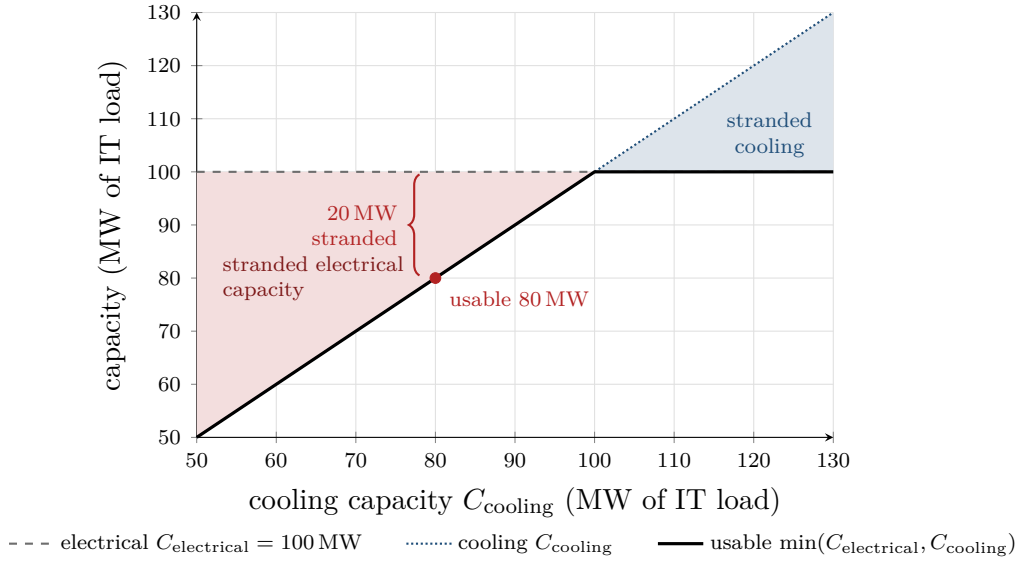


Figure 11.8: Usable IT capacity as the minimum of an electrical limit of 100 MW and a cooling limit C_{cooling}, illustrative (equation (11.18)). At $C_{\text{cooling}} = 80$ MW the facility can host 80 MW and 20 MW of electrical capacity is stranded.

11.6.3 Spatial metrics

PUE is one number for the whole facility and has no spatial resolution. A single high-density rack that needs excessive local fan power, from in-row coolers running at full speed or from its own server fans racing against a high inlet temperature, barely moves the facility figure; the 30 kW rack of Section 11.4.3 is invisible at the third decimal. The problem is real and local, and only a local metric shows it.

Finer metrics divide the same balance by zone or by rack. A local cooling ratio for zone or rack i,

$$\frac{P_{\text{cool},i}}{P_{\text{IT},i}}, \tag{11.19}$$

compares the cooling power spent on that zone with the IT power it hosts; zone thermal indicators compare inlet temperatures with their limit, and airflow indicators compare the air delivered with the air the racks draw. With illustrative numbers, a facility whose zones average 0.15 in equation (11.19) and whose one high-density row runs at 0.40 has a row that costs more than twice as much to cool per kilowatt as the rest, a fact that the facility PUE cannot show. These local quantities are rarely metered directly: the cooling power attributable to a zone is usually calculated from fan speeds and coil duties, or inferred from the air and water balances.

11.6.4 The ratio as evidence about the system

The previous subsections treat PUE as a performance figure. It is more useful as an observation of an underlying physical system. Written out, the PUE is a function of the load, the weather, the parameters of each component and the operating state,

$$\text{PUE} = f\big(P_{\text{IT}},\ T_{\text{amb}},\ \theta_{\text{chiller}},\ \theta_{\text{fan}},\ \text{control state},\ \text{equipment availability}\big), \tag{11.20}$$

where θ_{chiller} collects the chiller parameters (its efficiency curve, its fouling state) and θ_{fan} the fan parameters. An anomalous PUE, a positive residual of equation (11.16), is then evidence

about which of these has changed. Each candidate cause is a hypothesis, and the question is the probability of each given the PUE and the other telemetry,

$$\Pr(H_i \mid \text{PUE}, \text{telemetry}) \propto \Pr(\text{PUE}, \text{telemetry} \mid H_i)\ \Pr(H_i), \tag{11.21}$$

with four illustrative hypotheses: H_1, chiller degradation; H_2, fan over-operation; H_3, pump inefficiency; H_4, bias in a power meter.

The hypotheses are separable because each predicts a different pattern in the other measurements. Table 11.2 lists the signature of each and works an illustrative update. Suppose the residual is +0.03, the chiller's electrical power per unit of heat removed is 8 % above its reference, fan speeds and pump powers are unchanged, and the sum of the submeters agrees with the main meter within 0.5 %. That combination is likely under H_1 and unlikely under the others, and with the prior and likelihood values of the table the posterior probability of chiller degradation is about 0.82. Measured quantities enter as data, the expected PUE is calculated, and the conclusion, a ranked list of causes with probabilities, is inferred.

Table 11.2: Four hypotheses for an anomalous PUE, their signatures in other telemetry, and an illustrative update for the observation in the text (residual +0.03, chiller power per unit heat +8 %, fans and pumps unchanged, meters consistent). Posterior is prior times likelihood, normalized.

Hypothesis	Signature in other telemetry	Prior	Likelihood	Posterior
H_1 chiller degradation	chiller power per unit heat removed rises at the same load and condenser temperature	0.30	0.60	0.82
H_2 fan over-operation	fan speeds or fan power rise at unchanged airflow demand	0.30	0.05	0.07
H_3 pump inefficiency	pump power rises at unchanged flow and pressure	0.20	0.10	0.09
H_4 meter bias	PUE shifts while the component submeters do not; sum of submeters disagrees with the main meter	0.20	0.02	0.02

11.6.5 Questions the ratio cannot answer, and a what-if analysis

The operating questions that matter are of a kind the ratio cannot answer: why facility energy is increasing; which subsystem is responsible; whether the change is consistent with the weather and the IT load or indicates a fault; whether an efficiency gain raises the thermal risk; and what happens if the chilled-water setpoint changes. Each requires the relations among the components: the chiller's dependence on the water temperatures, the coil's dependence on the water and air flows, the racks' dependence on the air delivered to them. A physical model of the facility, simulated forward and conditioned on its measurements, can answer them. The ratio, a quotient of two meters, cannot.

The last question is a what-if analysis. Suppose the chilled-water supply setpoint $T_{\text{CHW,s}}$ is raised from 7°C to 10°C. A facility model predicts three consequences, with illustrative values:

- the chiller power falls by 12 %, because a warmer evaporator narrows the temperature lift the compressor must provide;
- the annual PUE falls from 1.22 to 1.18;

- the probability that at least one rack exceeds its inlet limit during the year rises from 0.2 % to 3.5 %, because warmer water through the same coils delivers warmer supply air, and the racks with the least margin cross the limit first.

The second figure is not the first figure converted. At $P_{\mathrm{IT}} = 10\,\mathrm{MW}$, a PUE change of 0.04 is 0.4 MW; a 12 % cut in the chiller power of Table 11.1, 1.2 MW, is only 0.144 MW, or 0.014 of PUE. The rest, about 0.026, comes from the economizer: a warmer chilled-water supply can be produced by the cooling-tower water on more days of the year, which moves the mode boundaries of Figure 11.3 to the right and adds economizer hours. Only a model that holds both mechanisms produces the combined number.

The decision is whether the energy saving is worth the added thermal risk. At \$80/ MWh, 0.4 MW is 3504 MWh and about \$280 000 a year; the risk rises by 3.3 percentage points. By the reasoning of Section 11.4.6, the change pays only if a thermal violation costs less than $\$280\,000/0.033 \approx \$8.5$ million, or if the risk is first reduced, for example by fixing the airflow of the few racks with the least margin before raising the setpoint. That is equation (11.12) applied to one decision, and it needs the model; the PUE change alone would have approved the setpoint without seeing the risk.

11.6.6 PUE in context: a dashboard and the equations to keep

A dashboard that reports PUE alone invites every misreading of this chapter. A more useful operating view places it among the quantities that explain it and the quantities it trades against: the PUE itself; the IT power P_{IT} and the cooling power P_{cool} in megawatts, so that a change in the ratio can be attributed; the thermal margin, the smallest difference $T_{\max} - T_{\mathrm{in},i}$ over all racks; the cooling capacity margin, $\dot{Q}_{\mathrm{avail}} - \dot{Q}_{\mathrm{req}}$, with and without the largest running unit; the probability of a thermal violation over the coming day; and the expected energy cost. PUE in that view is one indicator among seven, read in the context of the others.

The equations to keep are four:

$$\mathrm{PUE} = \frac{P_{\mathrm{facility}}}{P_{\mathrm{IT}}}, \qquad P_{\mathrm{facility}} = P_{\mathrm{IT}} + P_{\text{non-IT}}, \qquad \mathrm{PUE} = 1 + \frac{P_{\text{non-IT}}}{P_{\mathrm{IT}}}, \tag{11.22}$$
$$r_{\mathrm{PUE}} = \mathrm{PUE}_{\mathrm{observed}} - \mathrm{PUE}_{\mathrm{expected}}.$$

The first defines the metric, the second is the balance it summarizes, the third isolates the overhead, and the fourth turns the metric into a diagnostic signal.

11.6.7 Metric and model

PUE answers one question: how much facility power is spent per unit of IT power. Table 11.3 sets that question beside the ones it does not answer, with the quantity or model that does answer each and the subsection where it is treated. The left column is short and the right is long, which is the summary of the chapter.

PUE is a useful metric, and it is not a model of the system. A metric summarizes a state; a model explains it, predicts it under conditions that have not occurred, and attaches probabilities to the risks a decision creates. The target of operating a data center is to minimize energy and cost while keeping thermal and reliability risk acceptable, which is equation (11.12), and PUE is one of its inputs, not its objective.

11.7 What to retain

Definition 11.1 (PUE as a metric)**.** Power usage effectiveness, $\mathrm{PUE} = P_{\mathrm{facility}}/P_{\mathrm{IT}} = 1 + P_{\text{non-IT}}/P_{\mathrm{IT}}$, measures the facility power spent per unit of IT power; it varies with load and

Table 11.3: What PUE answers and what it does not, with what answers each question instead.

Question	Answered by	Subsection
Answered by PUE		
How much facility power per unit of IT power?	$\text{PUE} = P_{\text{facility}}/P_{\text{IT}}$	11.1.1
Not answered by PUE		
Is the computing efficient?	useful work per unit of IT energy	11.2.1, 11.2.6
Is total energy falling?	P_{facility} and E_{facility} themselves	11.2.2
Is water use acceptable?	water per unit of IT energy or workload	11.2.4
Are emissions low?	CO_2e per unit of IT energy, by hour	11.2.5
Which subsystem is responsible?	component decomposition	11.4.2
Are racks overheating?	rack inlet temperatures, local metrics	11.4.3, 11.6.3
Is the system reliable, with adequate capacity under failure?	capacity margin with the largest unit lost	11.4.4
Is the facility near a dangerous boundary?	probability of a thermal violation	11.4.6
Is a change a fault or the weather?	residual against an expected PUE	11.5.4
How much IT load can the site host?	minimum over subsystem capacities	11.6.2

weather without any fault and says nothing about computing efficiency, water, carbon, thermal safety, resilience or capacity. It is useful as a summary and, conditioned on load and weather, as a residual signal, but the operating objective is to minimize energy and cost while keeping thermal and reliability risk acceptable.

Table 11.4 lists the chapter's objects and the subsection that defines each.

11.8 Exercises

Exercise 11.1. A facility's cooling power follows $P_{\text{cool}} = 0.8 + 0.12\,P_{\text{IT}}$ (MW), and all other non-IT power is negligible. (a) Compute the PUE at $P_{\text{IT}} = 4$, 8 and 16 MW, and the floor that PUE approaches at high load. (b) At what IT load does the PUE equal 1.25? (c) The IT load rises from 8 to 16 MW. Use equation (11.4) with the values at 8 MW to estimate the change in PUE, compare it with the exact change, and state which of the two powers caused it.

Exercise 11.2. A facility runs for 2000 h at $P_{\text{IT}} = 15$ MW with PUE = 1.30, 4000 h at $P_{\text{IT}} = 12$ MW with PUE = 1.18, and 2760 h at $P_{\text{IT}} = 5$ MW with PUE = 1.45. (a) Compute the simple average of the three PUE values weighted by hours, and the annual PUE of equation (11.14). (b) Explain which is larger and why. (c) In the first period the overhead splits as chillers 2.4 MW, fans 0.9 MW, pumps 0.6 MW, power-path losses 0.6 MW; compute each component's contribution to PUE and say which subsystem offers the largest opportunity.

Exercise 11.3. A site has a 60 MW grid connection. Its peak PUE is 1.30. (a) Compute the maximum IT load the connection supports. (b) An upgrade lowers the peak PUE to 1.20; compute the IT capacity released. (c) The cooling plant after the upgrade supports 47 MW of IT load and the UPS system 52 MW. Compute the usable capacity by equation (11.18), name the limiting subsystem, and compute the stranded capacity of each of the other two limits.

Table 11.4: The objects of the chapter and where each is defined.

Object	Expression	Role	Defined in
PUE	$P_{\text{facility}}/P_{\text{IT}}$, equation (11.1)	facility power per unit of IT power	11.1.1
Overhead form	$1 + P_{\text{non-IT}}/P_{\text{IT}}$	isolates the non-IT power	11.1.1
Change of a ratio	equation (11.4)	attributes a change to overhead or IT load	11.2.3
Workload efficiency	equation (11.5)	useful work per facility energy	11.2.6
Load dependence	$1 + k + P_0/P_{\text{IT}}$	poor PUE at low utilization	11.3.1
Weather dependence	equation (11.8)	daily and seasonal variation	11.3.3
Component overhead	equation (11.9)	which subsystem is responsible	11.4.1
Risk-priced objective	equation (11.12)	energy against thermal and reliability risk	11.4.7
Annual PUE	equation (11.14)	IT-weighted average of instantaneous PUE	11.5.1
Residual	r_{PUE}, equation (11.16)	detects degradation under same conditions	11.5.4
Capacity limit	$P_{\text{grid,max}}/\text{PUE}$	IT load a connection supports	11.6.1
Usable capacity	equation (11.18)	minimum over subsystems; stranded rest	11.6.2
Hypotheses H_1–H_4	Table 11.2	causes of an anomalous PUE	11.6.4

Exercise 11.4. A setpoint change at a facility with $P_{\text{IT}} = 20\,\text{MW}$ is predicted to lower the annual PUE by 0.03 and to raise the annual probability of a thermal violation from 0.5 % to 2.5 %. Electricity costs \$70/ MWh. (a) Compute the annual energy and cost saving. (b) Compute the violation cost above which the change is not worth making. (c) If a violation is estimated to cost \$15 million, should the change be made under equation (11.12), with only the energy and thermal terms? (d) A week after the change the observed PUE is 1.21 while a model of the facility at the week's load and weather predicts 1.19, with a residual standard deviation of 0.005. Compute the residual and its size in standard deviations, and name two hypotheses that would explain it.

11.9 Connection to the TERRA framework

PUE is a quotient of two edge flows in the facility's electrical network. The framework treats it as one calculated observation of a system whose structure the earlier chapters built, and the chapter's objects map onto that structure directly.

- **The facility power balance is a conservation node.** Equation (11.2) is the power balance of the facility's main electrical bus: the utility edge flows in, and the IT edge and the edges to chillers, fans, pumps and the power-path losses flow out. PUE is a ratio of two measured edge flows; the component decomposition of Table 11.1 is the same node with every outgoing edge written separately, and the mismatch between the submeters and the main meter is a residual of that balance, the evidence that separates H_4 from the physical hypotheses.
- **The expected PUE is a forward simulation.** The electrical, air and water mass, and thermal networks are coupled through components with parameters: the chiller's

efficiency as a function of $T_{\mathrm{CHW,s}}$ and the condenser temperature, the fan's cubic law, the pump's affinity law, the UPS efficiency η, the economizer's mode boundaries. Simulating them at the measured P_{IT} and weather gives $\mathrm{PUE}_{\mathrm{expected}}$, and r_{PUE} of equation (11.16) is the residual between that simulation and the meters.

- **Hypotheses are parameter shifts with a posterior.** H_1 to H_3 of Table 11.2 are changes in $\theta_{\mathrm{chiller}}$, θ_{fan} and the pump efficiency; H_4 is a bias on a measurement edge. Equation (11.21) is the ranked posterior over them, and the next measurement, for example a temporary flow meter on the condenser loop, is chosen for how much it would separate the leading hypotheses.

- **Risk is a probability computed on the network.** The violation probabilities of Sections 11.4.6 and 11.6.5 are the probability that the cooling available at some rack node is below the cooling it requires, given the posterior over unmeasured quantities such as rack airflow and the recirculation fraction α. They enter equation (11.12) as the priced risk terms.

- **Capacity is a minimum over component bounds.** Equation (11.18) is the largest IT load for which every component stays within its bounds, and stranded capacity is the slack of the components that do not bind. Evaluated at rack nodes rather than for the facility, the same bounds give the spatial metrics of Section 11.6.3.

In summary, PUE is the facility power per unit of IT power, a single ratio that varies with load and weather, hides where the overhead is spent and what the efficiency costs in risk, and becomes informative when it is read as an observation of a physical system: conditioned on its drivers it is a residual, decomposed it points to a subsystem, and set beside a computed probability of violation it becomes one term of an operating decision. Redundancy and reliability, which turn the loss of a chiller under a lean operating policy into a probability, follow in a later chapter.

Chapter 12

Redundancy and Reliability

12.1 Redundancy levels and layers

The chapters so far described a data center in which every component works: power flows from the utility to the racks, heat flows from the racks to the outdoor air, and controls keep both in balance. Components do fail, and they are also taken out of service on purpose for maintenance. A data center is built so that neither event interrupts the critical IT service. Installing spare equipment is the usual means, but spare equipment is not the same thing as reliability. The question of this chapter is

how a data center is designed and evaluated so that equipment failures do not interrupt critical IT service, and what decides whether a given design tolerates a given failure at a given moment.

The chapter answers it in five steps. It first names the redundancy levels and the layers at which they apply. It then shows how shared dependencies undermine redundancy, and how reliability is computed for components in series and in parallel. It defines availability and the role of maintenance and of degraded capacity. It treats contingency analysis as a problem in time, with event trees and fault trees as the tools for sequences and combinations of failures. Finally, it separates the reliability of a design on paper from the risk of the facility as it operates now.

Jargon: redundancy, N, N+1, N+2, 2N, single point of failure, common-cause failure, contingency, concurrent maintainability, fault tolerance, MTBF, MTTR

Redundancy is equipment installed beyond what the load requires, so that the load can still be carried when some equipment is unavailable. N is the number of units required to carry the design load. An N design installs exactly N; an $N{+}1$ design installs one spare unit; an $N{+}2$ design installs two; a $2N$ design installs two complete and separate systems, each able to carry the whole load. A *single point of failure* is a component whose failure alone interrupts the critical service. A *common-cause failure* is the failure of several components from one shared cause. A *contingency* is a specified failure or outage whose consequences are analyzed, such as the loss of one chiller. *Concurrent maintainability* is the ability to take any component out of service for planned maintenance without interrupting the critical load. *Fault tolerance* is the ability to continue serving the critical load, without interruption, after an unplanned failure. *MTBF* (mean time between failures) is the average operating time from one failure to the next; *MTTR* (mean time to repair) is the average time to restore a failed component to service.

Key idea

A redundancy label such as $N{+}1$ or $2N$ describes an architecture, not the risk the facility carries now. That risk depends on the topology, the state and capacity of each component, the causes the components share, the load, the weather and the speed of the response, and the operational question is which failures the system, as it exists at this moment, can actually tolerate.

This section defines the four redundancy levels with one running example, a chiller plant that must remove 6 MW of heat, and then lists the layers of the facility at which redundancy appears.

12.1.1 The N design

A design at N installs exactly enough equipment to carry the required load and nothing more. For a required cooling of $\dot{Q}_{\text{req}} = 6\,\text{MW}$ provided by chillers rated 2 MW each, the number required is

$$N = \left\lceil \frac{\dot{Q}_{\text{req}}}{\dot{Q}_{\text{unit}}} \right\rceil = \left\lceil \frac{6\,\text{MW}}{2\,\text{MW}} \right\rceil = 3, \tag{12.1}$$

where $\dot{Q}_{\text{unit}}$ is the rated capacity of one unit and $\lceil \cdot \rceil$ rounds up to the next integer. Three chillers give exactly 6 MW. If one fails, the remaining two give

$$\dot{Q}_{\text{avail}} = 2 \times 2\,\text{MW} = 4\,\text{MW} < 6\,\text{MW} = \dot{Q}_{\text{req}},$$

a shortfall of 2 MW, one third of the load. An N design has no spare: every failure, and every unit taken out for maintenance, reduces the available capacity below the requirement. N designs are common for equipment whose loss is tolerable or whose repair is fast, and they are the reference against which the other levels are named.

12.1.2 N+1 and N+2

An $N{+}1$ design installs one unit more than required. For the same plant, four chillers give 8 MW installed against 6 MW required; if any one fails, the remaining three give

$$\dot{Q}_{\text{avail}} = 3 \times 2\,\text{MW} = 6\,\text{MW} = \dot{Q}_{\text{req}},$$

exactly enough. The design intent of $N{+}1$ is that the failure of any single component does not interrupt the required capacity. A second failure while the first is unrepaired leaves 4 MW, short by 2 MW.

An $N{+}2$ design installs two spares: five chillers, 10 MW installed, of which any two may be unavailable at once and the remaining three still give 6 MW. The usual reason is to tolerate one failure while another unit is out for maintenance (Section 12.3.1). The added resilience has a price in capital cost, in maintenance of the extra unit, and in plant-room space and piping.

Figure 12.1 compares the levels. The horizontal axis is the number of chillers out of service, for any reason; the vertical axis is the capacity that remains, in megawatts. Each line is one design: N with three chillers, $N{+}1$ with four, $N{+}2$ with five, and $2N$ with two systems of three, drawn on the assumption that the two systems can share the load. The dashed red line is the requirement. A design tolerates as many units out as the point at which its line reaches the requirement: none for N, one for $N{+}1$, two for $N{+}2$ and three for $2N$ under the pooling assumption. Every unit out removes the same 2 MW, so the lines are parallel and the redundancy level only shifts them upward.

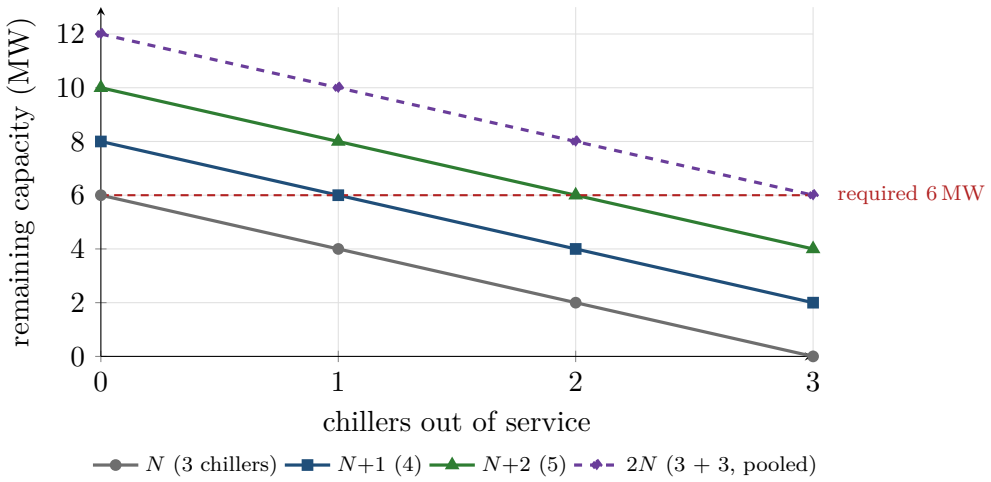


Figure 12.1: Capacity remaining as chillers go out of service, for the redundancy levels of a plant with $\dot{Q}_{\text{req}} = 6\,\text{MW}$ and 2 MW chillers. A design tolerates the number of outages at which its line meets the requirement. The $2N$ line assumes that the two systems can share the load; if they cannot, $2N$ tolerates the loss of one whole system, not any three units.

12.1.3 The 2N design

A $2N$ design duplicates the full required capacity. The plant becomes two separate systems of three chillers each, system A with 6 MW and system B with 6 MW, and either one alone supports the whole load. The installed capacity is 12 MW, twice that of N and one and a half times that of N+1.

The extra cost buys protection against a different kind of failure. An N+1 plant protects against the loss of any one chiller inside the plant, but its four chillers usually share a header, a set of pumps, a switchboard and a control system, and the loss of any of those removes all four at once. A $2N$ design protects against the loss of an entire system, including its shared elements, provided the two systems share nothing. That proviso is the whole argument: $2N$ is more robust than N+1 only if the two paths are genuinely independent. Two systems that share a switchboard, a water supply or a control network are, with respect to that element, a single system. Section 12.2.3 puts numbers on this comparison and shows that, with independent units and no shared element, $2N$ arranged as two separate systems can even be less reliable than N+1; its advantage appears only when shared elements are counted.

12.1.4 Redundancy at several layers

Redundancy is applied at every layer of the facility, not only in the chiller plant. Table 12.1 lists the layers, the redundant elements usually installed in each, and an example of an element the redundant units commonly share. The electrical layer has two utility feeds A and B, UPS systems A and B, and power distribution units (PDUs) A and B, feeding the two power supplies of a dual-corded server (Chapter 3). The cooling layer has chillers 1 to 4, pumps A and B, and more computer room air handlers (CRAHs) than the load needs. The IT layer has dual power supplies in each server, redundant network switches, and workloads replicated across servers or across sites. The control layer has redundant controllers and redundant communication links.

The layers depend on one another. The UPS rooms need cooling, the chillers need electrical power, the controllers need both, and the servers need all three. A server is supplied only if every layer delivers to it at the same time, so reliability is an end-to-end property of the chain from the utility and the outdoor air to the server, not a property of any one layer. A $2N$ electrical design feeding a plant with a single chilled-water header is, end to end, no better than

Table 12.1: Redundancy by layer, with an example of a shared element that can defeat it. The examples are typical, not exhaustive.

Layer	Redundant elements	Example of a shared element
Electrical	utility feeds A and B; UPS A and B; PDUs A and B; generators $N+1$	one switchboard upstream of both UPS systems; one fuel tank for all generators
Cooling	chillers 1 to 4; pumps A and B; CRAH units $N+1$	one chilled-water header; one make-up water supply; one electrical feeder to the plant
IT	dual power supplies; redundant network switches; replicated servers and storage	one rack-level breaker feeding both cords; one storage array behind both replicas
Control	redundant controllers; redundant communication links	one control network or one configuration server for both controllers

that header.

12.2 Dependence and single points of failure

Redundancy protects only against failures that leave at least one path intact. This section identifies the components whose loss leaves none, draws the facility as a dependency graph, computes the reliability of components in series and in parallel, and shows how a shared cause breaks the independence on which the parallel formula rests.

12.2.1 Single points of failure

A single point of failure is a component whose failure alone can interrupt the critical service. In a design without redundancy every component is one. In a redundant design they hide in the elements that the redundant paths share. Typical examples are two UPS systems in a $2N$ arrangement that are both fed from one common switchboard, so that a fault on the switchboard removes both; and two cooling loops that both depend on one control network, one water supply, one common header or one electrical feeder. In each case the redundancy on the drawing is real for the units it duplicates and absent for the element it shares.

Shared infrastructure is rarely installed by mistake. It saves space, cost and complexity, it may be required by the utility connection or by the site, and it is often added later, by a modification that joins two systems for convenience. Finding single points of failure is therefore not a one-time design review but a recurring check of the facility as built and as modified, and it requires a description of what depends on what.

12.2.2 The dependency graph

That description is a dependency graph: each node is a component, and each arrow points from a component to the components that need it to function. Two chains reach every rack. The electrical chain runs from the grid through the switchgear, the UPS and the PDU to the server; the cooling chain runs from the chiller through the pump and the CRAH to the rack. A rack is served only if both chains deliver.

Figure 12.2 draws the two chains for an illustrative facility. The upper rows are the electrical chain, with the A and B paths drawn separately; the lower row is the cooling chain, with the

redundant units collapsed into one box per stage. Arrows point in the direction of dependence. Red boxes are elements that the redundant paths share: a switchboard common to both UPS systems, a chilled-water header common to all chillers, and a control network common to the chillers and the pumps (dashed, because it carries signals rather than power or water). Each red box is a single point of failure: removing it disconnects the rack from every source on its chain, however many redundant units sit upstream or downstream of it. In the language of graphs, it is a cut vertex between the sources and the rack. An unprotected single component on a critical path weakens the nameplate architecture down to the reliability of that component.

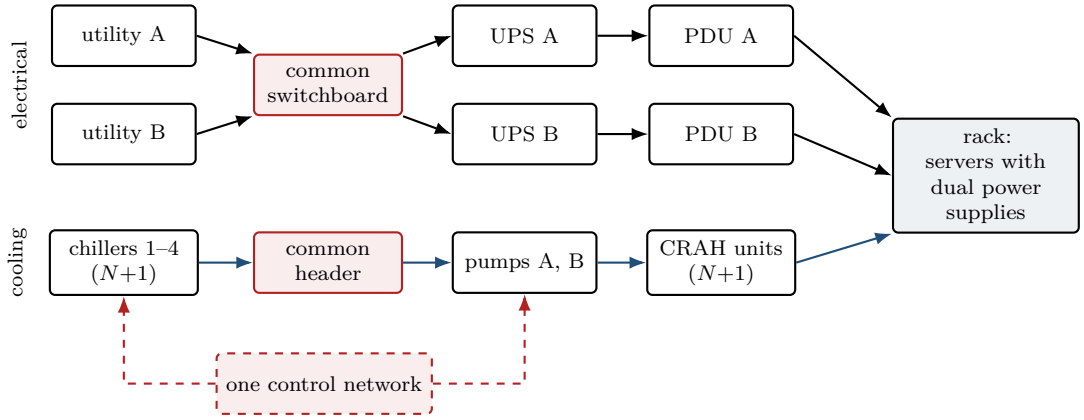


Figure 12.2: Dependency graph of an illustrative facility. Arrows point from a component to the components that need it. The electrical chain (top) has A and B paths; the cooling chain (bottom) has redundant units at each stage. Red boxes are shared elements and single points of failure: a switchboard common to both UPS systems, a header common to all chillers, and a control network (dashed) common to chillers and pumps.

12.2.3 Series and parallel reliability

The reliability R_i of a component is the probability that it operates without failure over a stated interval, such as one year or the duration of one outage. Two arrangements of components cover most of a dependency graph.

Series. When every one of the components A, B and C is required, as in a chain from a switchboard through a UPS to a PDU, the system works only if all of them work. If their failures are independent, the probability that all work is the product,

$$R_{\text{series}} = R_A\, R_B\, R_C = \prod_i R_i. \tag{12.2}$$

For three components of reliability 0.99 each, $R_{\text{series}} = 0.99^3 = 0.970299 \approx 0.9703$. The chain is less reliable than any of its links, and every component added in series lowers it further: the chain fails with probability 0.0297, almost three times the failure probability of one link. For small failure probabilities, $1 - R_{\text{series}} \approx \sum_i (1 - R_i)$: failure probabilities in series approximately add.

Parallel. When either A or B suffices, as with the two power supplies of a dual-corded server, the system fails only if both fail. If their failures are independent,

$$R_{\text{parallel}} = 1 - (1 - R_A)(1 - R_B) = 1 - \prod_i (1 - R_i). \tag{12.3}$$

With $R_A = R_B = 0.99$, $R_{\text{parallel}} = 1 - 0.01 \times 0.01 = 0.9999$: the failure probability falls from 10^{-2} to 10^{-4}, a hundredfold. Parallel failure probabilities multiply, which is the whole appeal of redundancy, and the multiplication is valid only if the failures are independent. Independence is the dangerous assumption.

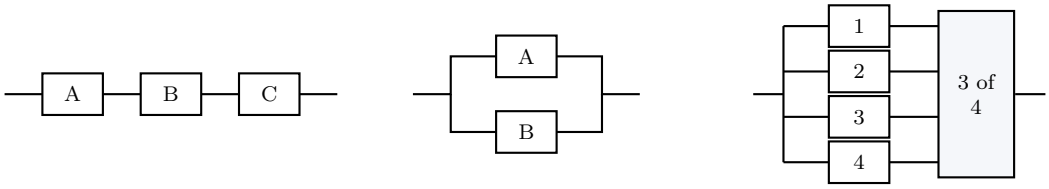
k of n. An N+1 or N+2 plant is neither pure series nor pure parallel: it works if at least $k = N$ of its n units work. With identical independent units of reliability R, the number that work is binomial, and

$$R_{k \text{ of } n} = \sum_{j=k}^{n} \binom{n}{j} R^j (1-R)^{n-j}. \tag{12.4}$$

For the N+1 chiller plant, three of four with $R = 0.99$,

$$R_{3 \text{ of } 4} = 0.99^4 + 4 \times 0.99^3 \times 0.01 = 0.960596 + 0.038812 = 0.999408.$$

Figure 12.3 draws the three arrangements as reliability block diagrams. In each panel the line runs from left to right, and the system works if a path of working blocks connects the two ends. Panel (a) is the series chain: one path through all three blocks. Panel (b) is the parallel pair: two paths, one through each block. Panel (c) is the N+1 plant: four blocks feeding a voting element that requires any three. The number under each panel is its reliability for blocks of reliability 0.99.



(a) series, all required $R = 0.99^3 = 0.9703$

(b) parallel, either suffices $R = 1 - 0.01^2 = 0.9999$

(c) N+1, any three of four $R = 0.999408$

Figure 12.3: Reliability block diagrams for blocks of reliability 0.99 with independent failures: (a) series, equation (12.2); (b) parallel, equation (12.3); (c) three of four, equation (12.4). The system works if working blocks connect the left end to the right end.

The formulas let the redundancy levels of Section 12.1 be compared on equal terms. Table 12.2 lists, for chillers of reliability 0.99 over an interval and independent failures, the probability that the plant carries 6 MW throughout. The first reliability column counts the chillers only. N is a series chain, $0.99^3 = 0.970299$. N+1 and N+2 are three of four and three of five. $2N$ is two separate systems of three, each a series chain of reliability 0.970299, in parallel: $1 - 0.029701^2 = 0.999118$. On the chillers alone, $2N$ arranged as two separate systems is *less* reliable than N+1, because each system needs all three of its own chillers while the N+1 plant can use any three of four.

The second reliability column adds one header of reliability 0.999: one header shared by the whole plant for N, N+1 and N+2, and one header per system for $2N$. The shared header sits in series with the whole plant and caps it at 0.999; N+1 falls to $0.999408 \times 0.999 = 0.998409$ and N+2 to 0.998990. The $2N$ systems each include their own header, $0.970299 \times 0.999 = 0.969329$, and in parallel give $1 - 0.030671^2 = 0.999059$, now the highest of the four. The ranking reverses as soon as shared elements are counted, which is what $2N$ is for.

Principle 12.1. *Failure probabilities in series approximately add and failure probabilities in parallel multiply, but only if the failures are independent. A component shared by redundant paths sits in series with all of them and caps the reliability of the whole at its own.*

Table 12.2: Probability that a chiller plant carries $\dot{Q}_{\mathrm{req}} = 6\,\mathrm{MW}$ over an interval, for chillers and headers of reliability 0.99 and 0.999 with independent failures (illustrative).

Design	Units	Success rule	Chillers only	With header(s)
N	3 chillers	all three	0.970299	0.969329 (one header)
$N{+}1$	4 chillers	any three of four	0.999408	0.998409 (one header)
$N{+}2$	5 chillers	any three of five	0.999990	0.998990 (one header)
$2N$	2 systems of 3	all three of either system	0.999118	0.999059 (one header per system)

12.2.4 Common-cause failures

Independence fails whenever two redundant components depend on the same thing. Suppose UPS A and UPS B are electrically separate but both sit in rooms cooled by the same cooling system C. If C fails, both UPS rooms overheat and both UPS systems shut down to protect themselves. In the dependency graph, C is a parent of both, $C \to A$ and $C \to B$, and in its presence the joint probability of failure is not the product of the marginal ones,

$$\Pr(A_f, B_f) \neq \Pr(A_f)\,\Pr(B_f), \tag{12.5}$$

where A_f and B_f denote the failure of A and of B. The joint probability can be far higher than the product.

A simple model shows by how much. Let the shared cause occur with probability q over the interval and fail both paths; otherwise let each path fail independently with probability p'. Hold each path's own failure probability at the observed 0.01, so that $q + (1-q)\,p' = 0.01$ and $p' = (0.01 - q)/(1-q)$. Then

$$\Pr(A_f, B_f) = q + (1-q)\,p'^2. \tag{12.6}$$

With $q = 0$ this is the independent value 10^{-4}. With $q = 0.001$, a tenth of each path's failure probability shared, $p' = 0.009/0.999 = 0.009009$ and

$$\Pr(A_f, B_f) = 0.001 + 0.999 \times 0.009009^2 = 0.001 + 0.000081 = 0.001081,$$

eleven times the independent value. The shared part dominates as soon as it exceeds the square of the independent part, which for failure probabilities of order 10^{-2} means as soon as it exceeds about 10^{-4}.

Figure 12.4 shows the structure and the effect. On the left is the graph: the shared cooling system C is a parent of both UPS paths. On the right, the horizontal axis is the probability q of the shared cause and the vertical axis, on a logarithmic scale, is the probability that both paths fail. The dashed gray line is the independent value 10^{-4}, the dotted line the failure probability 10^{-2} of one path alone, and the red curve equation (12.6). The curve leaves the independent value at once and rises almost in proportion to q; at $q = 0.01$, when all of each path's failure probability is shared, the pair is exactly as reliable as one path, and the redundancy adds nothing. The same effect for generators sharing a fuel line was treated in Chapter 4; it applies to every redundant pair in Table 12.1.

Principle 12.2. *Redundancy does not imply independence. A shared parent cause makes the failures of redundant paths dependent, and the joint failure probability is then set by the shared part, not by the product of the individual failure probabilities.*

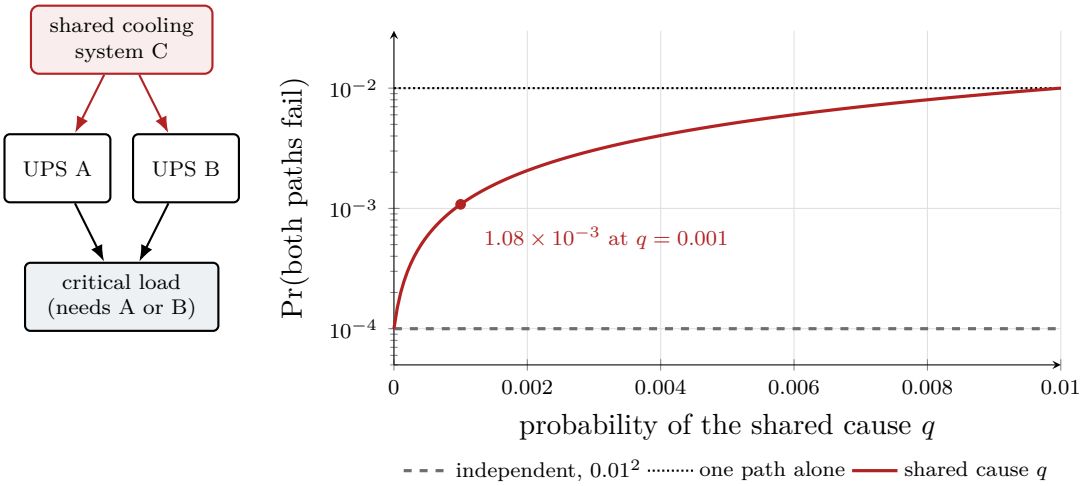


Figure 12.4: Common-cause failure of a redundant pair, illustrative. Left: a shared cooling system is a parent of both UPS paths. Right: probability that both paths fail against the probability q of the shared cause, $q + (1-q)p'^2$ with $p' = (0.01 - q)/(1 - q)$ (equation (12.6)), each path's own failure probability held at 0.01. A shared cause of probability 0.001 multiplies the joint failure probability by eleven.

12.3 Availability, maintenance and capacity states

Reliability counts failures over an interval; operation also involves planned outages, repair times and units that run at reduced capacity. This section treats maintenance as a temporary loss of redundancy, distinguishes concurrent maintainability from fault tolerance, defines availability and its expression in nines, separates it from reliability, and replaces the binary up-or-down state of a unit by its available capacity.

12.3.1 Maintenance and concurrent maintainability

Equipment must be serviced: chillers have their tubes cleaned and compressors overhauled, UPS systems have their batteries tested and replaced, breakers are inspected. A unit taken out for planned maintenance is unavailable exactly as if it had failed. An $N{+}1$ facility with one unit out for maintenance is temporarily an N facility, and a second, unplanned failure during the maintenance window gives $\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}}$.

An illustrative calculation shows how much the risk changes. Let each chiller of the $N{+}1$ plant fail at a constant rate $\lambda = 1/\text{MTBF} = 10^{-4}$ per hour (an MTBF of 10 000 h) and take $\text{MTTR} = 10$ h to repair. With all four chillers in service, a shortfall requires two to be down at once: one fails, at a rate 4λ, and a second fails during the first one's repair, with probability about $3\lambda \times \text{MTTR}$. The shortfall rate is about

$$4\lambda \times 3\lambda\,\text{MTTR} = 4 \times 10^{-4} \times 3 \times 10^{-4} \times 10 = 1.2 \times 10^{-6} \text{ per hour},$$

or about 5.8×10^{-5} over a 48 h interval. With one chiller out for a 48 h overhaul, any failure of the other three causes a shortfall, with probability

$$1 - e^{-3\lambda \times 48} = 1 - e^{-0.0144} = 0.0143$$

over the same 48 h, about 250 times higher. Risk is therefore not a fixed property of the design: it changes from hour to hour with the maintenance state.

Concurrent maintainability is the property that removes this exposure from planned work: every component can be taken out of service for maintenance without interrupting critical

operation. It requires that the topology and the capacity support the load with any one component intentionally unavailable, including the shared elements of Section 12.2.1. The $N+1$ chiller plant of Figure 12.2 is not concurrently maintainable, because its single header cannot be serviced without stopping all chillers; a second header with isolation valves would make it so. Concurrent maintainability is stricter than surviving random failures in one respect: it requires that every element, including pipes, valves, switchboards and busbars, has an alternative path, not only the machines. A design that is concurrently maintainable still runs at N during the maintenance itself, so tolerating a failure during maintenance requires $N+2$ or $2N$.

12.3.2 Fault tolerance

Fault tolerance is the ability to continue operating after an unplanned failure without interrupting the critical load. The standard example is a $2N$ UPS arrangement: UPS A fails, and UPS B carries the whole IT load without the servers noticing. That outcome requires more than a second UPS. It requires spare capacity, since UPS B must be rated for the full load and each UPS of a $2N$ pair therefore normally runs at half load or less; a suitable topology, with dual-corded servers or static transfer switches for single-corded ones (Chapter 3); automatic transfer, because no operator can act within the few milliseconds the servers' power supplies can ride through; correct controls and protection settings, so that a fault on path A does not trip path B; and no shared failure mode between the two paths.

Concurrent maintainability and fault tolerance are different properties. A concurrently maintainable design may rely on manual, planned switching to move a load from one path to the other before maintenance, which is acceptable for planned work and useless for a sudden failure. A fault-tolerant design must do the same transfer automatically and at once. Industry classifications of data center designs, such as the Uptime Institute's tiers, grade them by these two properties, with concurrent maintainability at one level and fault tolerance at the next. Both are properties of the whole system, not of the count of its units.

12.3.3 Availability, MTBF and MTTR

A repairable unit alternates between up intervals, which end in a failure, and down intervals, which end in a repair. Its availability A is the long-run fraction of time it is up,

$$A = \frac{\text{uptime}}{\text{uptime} + \text{downtime}} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}, \tag{12.7}$$

where the second form follows because each cycle contributes on average one MTBF of uptime and one MTTR of downtime. High availability requires failures that are infrequent, a long MTBF, and recovery that is fast, a short MTTR. The unavailability $1-A = \text{MTTR}/(\text{MTBF}+\text{MTTR})$ is approximately MTTR/MTBF when repairs are short compared with the time between failures.

For a unit with MTBF $= 10\,000$ h and MTTR $= 10$ h,

$$A = \frac{10\,000}{10\,000 + 10} = \frac{10\,000}{10\,010} = 0.999001 \approx 0.999.$$

Over a year of 8760 h, the unit fails on average $8760/10\,010 = 0.875$ times and is down $0.875 \times 10 = 8.75$ h, the same as $(1 - A) \times 8760$ h.

System availability differs from unit availability through redundancy and dependencies, by the same rules as reliability. Two such units in parallel, with independent failures and independent repairs, are both down a fraction $(1 - A)^2 \approx 10^{-6}$ of the time, an availability of about 0.999999; two in series are up a fraction $A^2 = 0.998$ of the time. The common causes of Section 12.2.4 raise the parallel figure in the same way as they raise the joint failure probability, and a shared repair crew or a shared spare part makes the repairs dependent as well.

12.3.4 The nines

Availability is quoted as a number of nines: 99 % is two nines, 99.999 % is five. The figures are easier to judge as downtime per year, $t_{\text{down}} = (1 - A) \times 8760\,\text{h}$. Table 12.3 works them out. Each row is one level of availability; the columns give the downtime per year in hours and in the most readable unit, and the longest single outage allowed if the facility has exactly one failure per year. Two nines allow about three and a half days of downtime per year; five nines allow about five minutes. Each added nine divides the allowed downtime by ten, and each is more expensive than the last, because it requires either more redundancy, which lengthens the MTBF of the system, or faster detection and automatic recovery, which shortens its MTTR. Five minutes per year is shorter than one manual repair, so five nines is reachable only by fault tolerance, never by fast repair alone.

Table 12.3: Availability and the downtime per year it allows, for a year of 8760 h. The last row is the example unit of Section 12.3.3.

Availability A	Downtime (h/yr)	Downtime	Longest outage, one failure per year
0.99 (two nines)	87.6	3.65 days	87.6 h
0.999 (three nines)	8.76	8.76 h	8.76 h
0.9999 (four nines)	0.876	52.6 min	52.6 min
0.99999 (five nines)	0.0876	5.26 min	5.26 min
0.999001 (MTBF 10 000 h, MTTR 10 h)	8.75	8.75 h	0.875 failures of 10 h each

12.3.5 Reliability versus availability

Reliability and availability answer different questions. Reliability is the probability of operating without any failure over an interval; availability is the probability of being operational at the moment the service is needed. For a unit with a constant failure rate $\lambda = 1/\text{MTBF}$, the reliability over an interval of length t is

$$R(t) = e^{-\lambda t} = e^{-t/\text{MTBF}}, \tag{12.8}$$

which depends on the MTBF only, while the availability of equation (12.7) depends on the ratio of MTTR to MTBF. A unit that fails often but is repaired very fast can have a high availability and a low reliability.

Figure 12.5 shows two illustrative units with the same availability. The horizontal axis is the length of the interval in hours; the vertical axis is the probability of no failure in that interval, equation (12.8). Unit X has an MTBF of 10 000 h and an MTTR of 10 h; unit Y has an MTBF of 500 h and an MTTR of 0.5 h. Both have $A = 0.999$. Over 1000 h, unit X runs without failure with probability $e^{-0.1} = 0.905$, unit Y with probability $e^{-2} = 0.135$. For a service whose cost is set by time down, the two are equivalent. For a service whose cost is set by each interruption, such as a long-running computation that must restart from its last checkpoint after every failure, unit Y is twenty times worse: it interrupts the service about 17.5 times per year against 0.875.

12.3.6 Capacity states and degraded N+1

The reliability and availability formulas treat each unit as either up or down. Real units have intermediate states. A chiller rated 2 MW may deliver only 1.5 MW because its condenser is

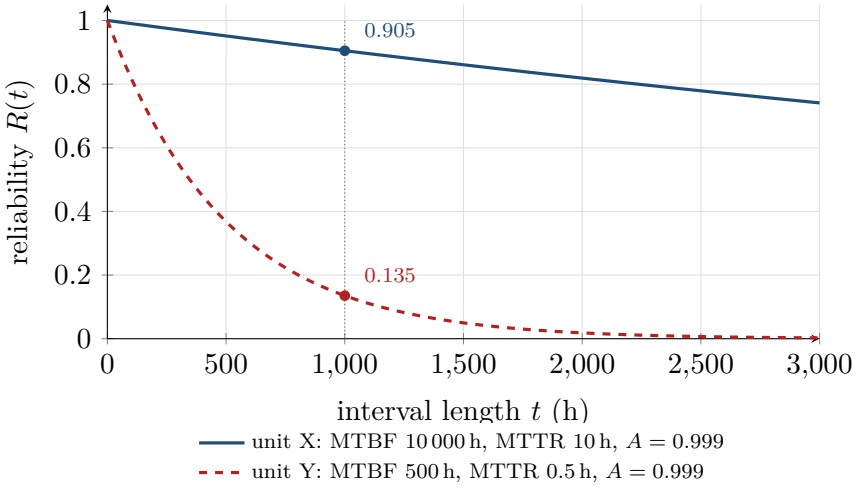


Figure 12.5: Reliability over an interval for two illustrative units of equal availability, $R(t) = e^{-t/\mathrm{MTBF}}$ (equation (12.8)), constant failure rates assumed. Unit Y fails twenty times as often as unit X and is repaired twenty times as fast, so both are up 99.9 % of the time, but over 1000 h unit X avoids failure with probability 0.905 and unit Y with 0.135.

fouled, its refrigerant charge is low, or the outdoor air is hot and its condensing temperature high (Chapter 7). The chiller is running, its status point says so, and the effective redundancy of the plant has nevertheless changed. The binary state is too crude; each unit has an available capacity $\dot{Q}_{\mathrm{avail},i}$ that depends on its condition and on the operating conditions, and the plant's available capacity is the sum over the units in service,

$$\dot{Q}_{\mathrm{avail}} = \sum_{i \in \text{in service}} \dot{Q}_{\mathrm{avail},i}. \tag{12.9}$$

Take the $N{+}1$ plant again: four chillers of 2 MW nominal, a requirement of 6 MW. On a hot day each chiller is derated to 1.6 MW. All four together give $4 \times 1.6 = 6.4$ MW, still above the requirement. With one chiller failed, the remaining three give $3 \times 1.6 = 4.8$ MW, short by 1.2 MW. The plant is labeled $N{+}1$ and on that day it cannot survive one failure: in capacity terms it is an N plant with 0.4 MW to spare.

Figure 12.6 shows how the condition depends on the weather. The horizontal axis is the outdoor temperature; the vertical axis is the plant capacity in megawatts. Each chiller is assumed to deliver its rated 2 MW up to 30°C and to lose 0.04 MW per kelvin above it, reaching 1.6 MW at 40°C; the assumption is illustrative. The upper line is four chillers in service; the lower line is three, after one failure; the dashed red line is the requirement, which coincides with the three-chiller line up to 30°C because the plant was sized with no margin beyond the spare. Above 30°C the three-chiller line falls below the requirement, and the shaded area is the shortfall that one failure would cause. A nominal $N{+}1$ plant at exact sizing loses its single-failure tolerance at the first kelvin of derating; the redundancy label hides that dependence on operating conditions.

Principle 12.3. *A unit is not up or down; it has an available capacity that depends on its condition and on the weather. A redundancy level stated in unit counts is a statement about nominal capacity and can be false at the operating point.*

12.4 Contingency analysis in time

Whether a facility tolerates a failure is decided by comparing capacity with requirement after the failure, and by how long the comparison can be unfavorable before a limit is crossed.

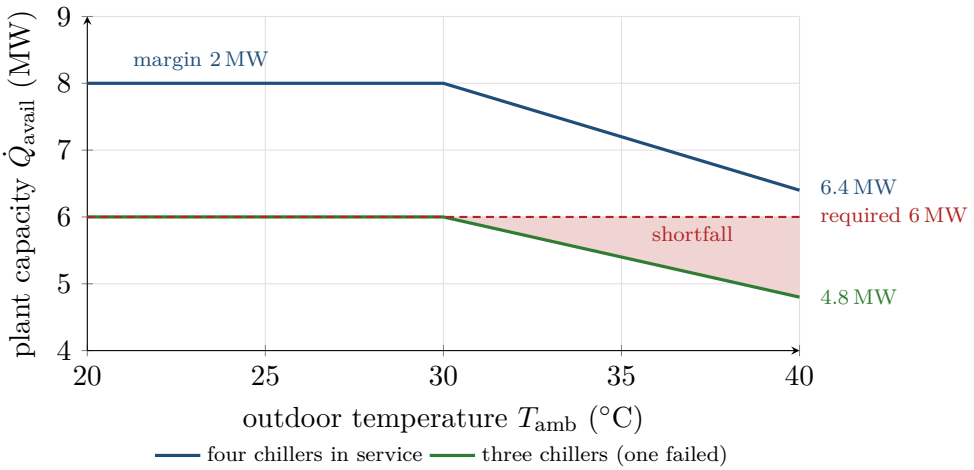


Figure 12.6: Capacity of a nominal $N{+}1$ plant of four 2 MW chillers against outdoor temperature, with an illustrative derating of 0.04 MW per chiller per kelvin above 30°C. Up to 30°C, three chillers exactly meet the 6 MW requirement; above it, one failure leaves a shortfall (shaded), 1.2 MW at 40°C.

This section defines the capacity margin and the contingency test, follows a failure across the electrical, thermal and IT domains, and computes the time to limit that bounds the recovery actions.

12.4.1 Capacity margin

The capacity margin is the difference between what is available and what is required,

$$M = \dot{Q}_{\text{avail}} - \dot{Q}_{\text{req}}, \tag{12.10}$$

the cooling form of $M = C - L$, capacity minus load. After a contingency c, such as the loss of one chiller, both terms change and the margin becomes

$$M_c = \dot{Q}_{\text{avail},c} - \dot{Q}_{\text{req},c}. \tag{12.11}$$

The requirement can change too: a lost pump changes the heat it adds to the water, and a failure that makes servers throttle reduces the heat they release. The facility is secure against a set of credible contingencies if

$$M_c \geq 0 \quad \text{for every credible contingency } c. \tag{12.12}$$

Checking this condition contingency by contingency is contingency analysis. The word “credible” carries the design choice: an $N{+}1$ facility is designed to satisfy equation (12.12) for the loss of any one unit, and a $2N$ facility for the loss of any one system.

12.4.2 Contingency screening and the N-1 analogy

Contingency analysis of a data center is the same procedure as the $N{-}1$ security analysis of a transmission system. There, each line or generator is removed in turn, the network is solved again, and the flows and voltages are tested against their limits; a network is $N{-}1$ secure if no single removal violates a limit. In a data center, each UPS, chiller, pump, switchboard or feed is removed in turn and the facility is tested on three questions: whether the electrical load is still served, whether the cooling is still adequate, and whether temperatures stay within their limits.

The $N-1$ of the power system and the $N+1$ of the data center describe the same goal from two sides: the first counts the elements that may be lost, the second the spare that allows it.

Table 12.4 screens six contingencies for an illustrative facility with a 6 MW IT load, a $2N$ UPS arrangement, four 2 MW chillers with two pumps of full capacity each, and a single chilled-water header. Chillers 1 and 2 are fed from switchboard A and chillers 3 and 4 from switchboard B. Each row is one contingency; the columns give whether the IT load stays powered, and the available cooling and the margin on a mild day and on a hot day when each chiller is derated to 1.6 MW. The UPS and pump contingencies pass in both weathers. The single-chiller contingency passes with zero margin on a mild day and fails on a hot one, as in Section 12.3.6. The loss of switchboard A is survived electrically, since UPS B carries the IT load, but it takes two chillers with it and leaves a cooling shortfall in any weather: an electrical contingency becomes a thermal one. The header is a single point of failure.

Table 12.4: Contingency screening of an illustrative facility: $\dot{Q}_{\text{req}} = 6$ MW, four 2 MW chillers (1.6 MW on a hot day), chillers 1 and 2 on switchboard A and 3 and 4 on switchboard B, a $2N$ UPS arrangement and a single chilled-water header. Margins in megawatts; negative margins fail the test of equation (12.12).

Contingency c	IT load powered	$\dot{Q}_{\text{avail},c}$, mild	M_c, mild	$\dot{Q}_{\text{avail},c}$, hot	M_c, hot
none (base case)	yes	8.0	+2.0	6.4	+0.4
one chiller	yes	6.0	0.0	4.8	−1.2
one pump (other at full flow)	yes	8.0	+2.0	6.4	+0.4
UPS A	yes, on UPS B	8.0	+2.0	6.4	+0.4
switchboard A (chillers 1, 2)	yes, on UPS B	4.0	−2.0	3.2	−2.8
chilled-water header	yes	0.0	−6.0	0.0	−6.0

12.4.3 Multi-physics contingencies

In a transmission system a contingency stays in one domain: a lost line redistributes electrical flows. In a data center a contingency crosses domains. A chiller failure reduces the cooling delivered; the temperature of the chilled water and then of the supply air rises; rack inlet temperatures rise; servers respond by raising their fan speeds, which raises the IT power, and, if the inlet keeps rising, by throttling their processors, which lowers it, or by shutting down. The chain runs

$$\begin{aligned}\text{chiller fails} \;&\rightarrow\; \dot{Q}_{\text{avail}} \text{ falls} \;\rightarrow\; \text{rack temperature rises} \\ &\rightarrow\; \text{servers throttle } (P_{\text{IT}} \text{ falls}) \text{ or shut down,}\end{aligned}$$

from the thermal domain into the IT domain and back into the electrical one. The switchboard row of Table 12.4 is the reverse chain, from the electrical domain into the thermal one. A contingency analysis that solves only the electrical network, or only the cooling plant, misses the propagation; the test of equation (12.12) has to be evaluated on the coupled electrical, air, water and thermal networks.

12.4.4 Time to limit

A negative margin does not cause an immediate violation. After a chiller failure, $\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}}$ at once, but the air, the water and the equipment store heat, and the rack inlet temperatures

may stay within limits for many minutes. The static question, whether $M_c \geq 0$, is therefore incomplete; the operational question is how long it takes, after the contingency, for a limit to be violated.

The heat balance of the room (Chapter 8) answers it. With an effective thermal capacitance C_{th} of the air, water and structure, a shortfall raises the temperature at the rate

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\dot{Q}_{\mathrm{req}} - \dot{Q}_{\mathrm{avail}}}{C_{\mathrm{th}}} = \frac{-M_c}{C_{\mathrm{th}}}, \tag{12.13}$$

and, extrapolating that rate linearly, the limit $T_{\max}$ is reached after

$$t_{\mathrm{limit}} = \frac{T_{\max} - T}{\mathrm{d}T/\mathrm{d}t}. \tag{12.14}$$

As an example, let $\dot{Q}_{\mathrm{req}} = 10\,\mathrm{MW}$ and let a chiller failure leave $\dot{Q}_{\mathrm{avail}} = 8\,\mathrm{MW}$, a shortfall of 2 MW. If the rack inlet temperature is observed to rise at 0.4°C per minute and the margin to the limit is 6°C, then

$$t_{\mathrm{limit}} = \frac{6^{\circ}\mathrm{C}}{0.4^{\circ}\mathrm{C/min}} = 15\,\mathrm{min}.$$

Equation (12.13) gives the effective capacitance implied by those two numbers: $C_{\mathrm{th}} = 2 \times 10^6\,\mathrm{W}/(0.4/60\,\mathrm{K/s}) = 3.0 \times 10^8\,\mathrm{J/K}$, that is, 300 MJ/K. The linear extrapolation is conservative when the rise is first order and slows as it goes (Chapter 8).

Recovery actions. The fifteen minutes are a window for recovery. Within it the operators or the controls can start a standby chiller, raise the chilled-water flow to use the remaining chillers' capacity fully, reduce the IT load, migrate workload to another hall or another site, or restore the failed equipment. Each action has its own duration and its own probability of success. The probability of an outage is therefore not the probability of the failure; it is the probability of the failure combined with the probability that no adequate response completes within the time to limit.

Figure 12.7 draws three illustrative responses to the chiller failure of the example. The horizontal axis is minutes after the failure; the vertical axis is the rack inlet temperature, starting at 21°C with a limit of 27°C (dashed red). With no recovery (gray) the temperature rises at 0.4°C/min and reaches the limit at 15 minutes. If a standby chiller of 2.5 MW comes on line at 6 minutes (green), the shortfall becomes a surplus of 0.5 MW and the temperature turns down from 23.4°C at 0.1°C/min. If the standby start fails and the operators cut the IT load to 80 % at 10 minutes (blue), the heat release falls to 8 MW, equal to the remaining capacity, and the temperature holds at 25°C, 2°C below the limit. The lines follow equation (12.13) with the capacitance of the example and assume that heat release equals IT power.

> **Principle 12.4.** *A contingency is survived or not according to whether an adequate response completes before the first limit is reached. The time to limit, $(T_{\max} - T)/(\mathrm{d}T/\mathrm{d}t)$, turns a negative capacity margin into a deadline, and the outage probability depends on the failure and on the response together.*

12.5 Event trees and fault trees

A failure and the responses to it unfold as a sequence of events, each of which can go more than one way, and an undesired outcome usually requires a combination of failures. This section describes the event tree, which follows a sequence forward from an initiating event, the fault tree, which decomposes an undesired event backward into its causes, and the probabilistic graph that connects reliability to diagnosis.

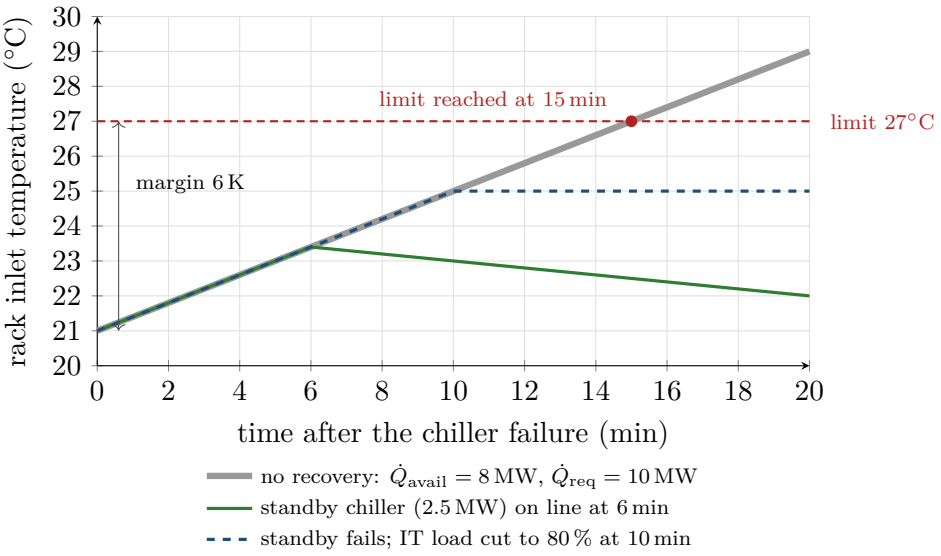


Figure 12.7: Rack inlet temperature after a chiller failure, illustrative, with $\mathrm{d}T/\mathrm{d}t = (\dot{Q}_{\text{req}} - \dot{Q}_{\text{avail}})/C_{\text{th}}$ (equation (12.13)) and $C_{\text{th}} = 300\,\text{MJ/K}$. Without recovery the 2 MW shortfall reaches the limit in 15 minutes; a standby chiller at 6 minutes reverses the rise; a 20 % IT load cut at 10 minutes stops it 2°C below the limit.

12.5.1 Sequences of events

The recovery of Section 12.4.4 is not one event but a sequence. Chiller 1 fails; the control system commands the standby chiller to start; the start succeeds or fails; the temperature rises while the outcome is pending; if the start fails, the operators shed IT load, if necessary, before the limit. Each step has its own probability, conditioned on the steps before it, and its own duration, which competes with the time to limit. The outcome is the result of a stochastic sequence, not a static probability attached to the design. Two facilities with identical equipment and different start logic, different operator procedures or different alarm thresholds have different outage probabilities.

12.5.2 Event trees

An event tree organizes such a sequence. It starts from an initiating event and, at each subsequent step, branches into the possible outcomes of that step, usually success and failure. Each path from the root to a leaf is one scenario, its probability is the product of the conditional probabilities along its branches, and the probabilities of all leaves sum to one.

Figure 12.8 is the event tree of a utility outage. The columns are the steps: the initiating event, whether the UPS carries the load, whether the generators start, and whether fuel is available for the duration. At each step the upper branch is success and the lower branch failure, labeled with illustrative conditional probabilities: the UPS carries the load with probability 0.999, the generators start with probability 0.99 given that the UPS has carried the load, and fuel is available with probability 0.995. The right-hand columns give the outcome of each path and its probability. The load is supplied throughout with probability

$$0.999 \times 0.99 \times 0.995 = 0.984065.$$

The other three paths are interruptions of different kinds: immediate, if the UPS fails (0.001); at battery depletion, if the generators fail to start ($0.999 \times 0.01 = 0.00999$); and when the fuel runs out, if it is unavailable ($0.999 \times 0.99 \times 0.005 = 0.004945$). The four probabilities sum to one, and the interruption probability given a utility outage is $1 - 0.984065 = 0.015935$.

Multiplied by the expected number of utility outages per year it gives an annual frequency of interruptions. The tree also shows which branch to improve: the generator start contributes almost two thirds of the total.

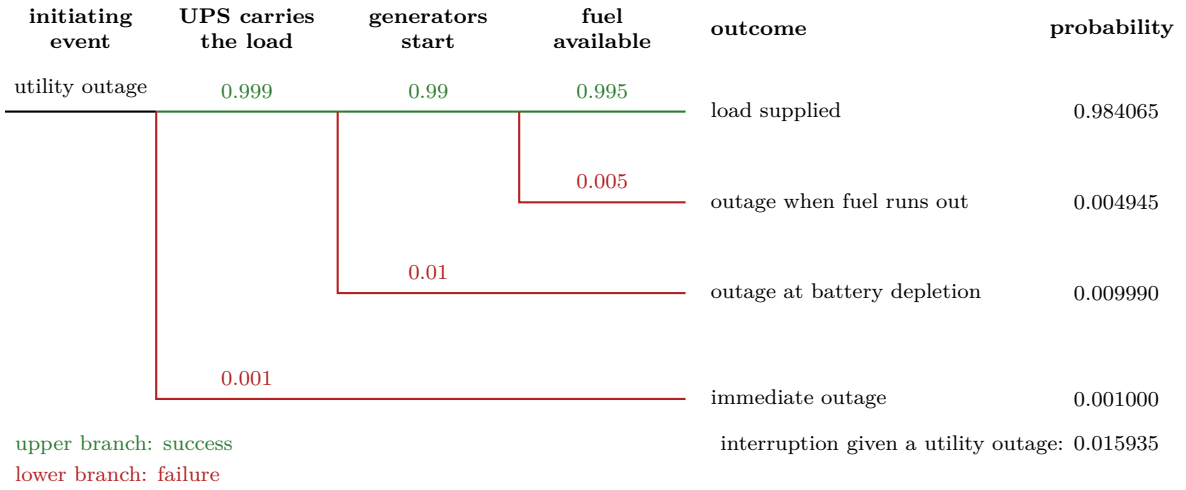


Figure 12.8: Event tree of a utility outage with illustrative conditional probabilities. Each path from the initiating event to an outcome is a scenario whose probability is the product of the branch probabilities; the four outcome probabilities sum to one. The generator plant is collapsed into one branch; a fuller tree would branch on each generator and on the transfer switch.

12.5.3 Fault trees

A fault tree works in the opposite direction. It starts from one undesired top event and decomposes it backward into the combinations of lower-level events that cause it, joined by logic gates: an AND gate when all of its inputs are needed to cause the output, an OR gate when any one suffices. For a dual-corded IT load, the top event "loss of IT power" occurs if path A fails AND path B fails; path A fails if UPS A fails OR PDU A fails OR the upstream breaker of path A fails; and similarly for path B. AND gates correspond to parallel redundancy and OR gates to series dependence, so the gates of a fault tree evaluate with equations (12.2) and (12.3) applied to failure probabilities.

Figure 12.9 draws the tree. The box at the top is the top event; gates are drawn with the usual symbols, a flat-bottomed dome for AND and a pointed shield for OR; circles are basic events, the component failures at which the decomposition stops, labeled with illustrative annual failure probabilities. The left branch is the combination just described. The right branch is a single basic event, the common switchboard of Figure 12.2, joined to the rest by an OR gate because it removes both paths by itself. With the probabilities shown, path A fails with probability

$$1-(1-0.005)(1-0.002)(1-0.001) = 1-0.992017 = 0.007983,$$

approximately the sum 0.008, and the same for path B. If the paths are independent, both fail with probability $0.007983^2 = 6.37\times10^{-5}$. Adding the switchboard, the top event has probability

$$1-(1-0.0005)(1-6.37\times10^{-5}) = 5.64\times10^{-4},$$

nine times the value without it. The switchboard, a single basic event with a probability of 0.0005, contributes almost nine tenths of the risk; the fault tree finds the single points of failure as the basic events that reach the top through OR gates only.

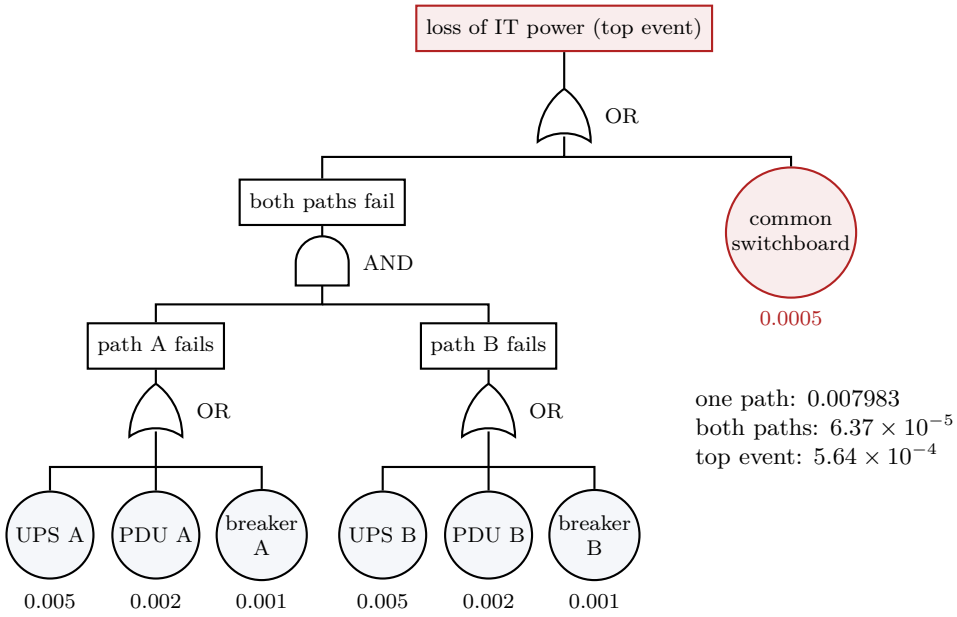


Figure 12.9: Fault tree for the loss of power to a dual-corded IT load, with illustrative annual failure probabilities of the basic events (circles). An AND gate requires all inputs, an OR gate any one. The common switchboard reaches the top event through OR gates only and contributes almost nine tenths of its probability.

12.5.4 The probabilistic-graph view

Event trees and fault trees are special cases of a more general object: a directed graph of random variables in which each variable depends on its parents, a Bayesian network. The common-cause graph of Figure 12.4 is one. The value of the general form is that it runs in both directions. Forward, it computes the probability of failures from the probabilities of their causes, as the trees do. Backward, it updates the probability of a cause from observations of its effects.

The standard example is fuel contamination, a common parent of three generators G1, G2 and G3 (Chapter 4). Let the prior probability of contamination be 0.01, and let a monthly test of a generator show rough running with probability 0.6 if the fuel is contaminated and 0.05 if it is not (illustrative). If G1's test shows rough running, Bayes' rule gives

$$\Pr(\text{contamination} \mid \text{G1 rough}) = \frac{0.01 \times 0.6}{0.01 \times 0.6 + 0.99 \times 0.05} = \frac{0.006}{0.0555} = 0.108,$$

eleven times the prior. Through the shared parent, the probability that G2 runs rough at its next test rises from $0.01 \times 0.6 + 0.99 \times 0.05 = 0.0555$ to $0.108 \times 0.6 + 0.892 \times 0.05 = 0.109$, and so does the probability that the whole generator plant fails in an outage. Generator test data thus update the belief in a hidden shared cause, and through it the reliability of the plant. That backward direction connects reliability analysis to probabilistic diagnosis, which Chapters 13 and 16 develop.

12.6 Operational risk

The tools so far compute what a design can tolerate. Operation needs the same answer for the facility as it is now, with its present load, weather, maintenance state, equipment condition and evidence. This section separates design reliability from operational reliability, defines expected risk and its attribution to components, adds the uncertainty of sensors and topology, and states reliability as partly a control problem with a supervisory question at its end.

12.6.1 Design and operational reliability

Design reliability is what the system should achieve on paper: the redundancy level of its drawings, the reliability computed from nameplate capacities and catalog failure rates. Operational reliability is what it can achieve now, given the present load, the condition of each unit, which units are out for maintenance, the weather, and the uncertainty in what the sensors report. The two can differ by a whole redundancy level: a facility designed as $N{+}1$ may currently operate as N because one unit is under maintenance (Section 12.3.1), because the weather has derated all units (Section 12.3.6), or because the load has grown into the spare.

Since loads, unit availability, weather, maintenance, topology and sensor evidence all change, the risk analysis has to be repeated as they change. Its output is then not the statement "this site is $N{+}1$" but a statement of the form: given current conditions, the probability of losing thermal margin under the most likely single contingency is 2.1 %. An illustrative calculation produces such a number. Suppose that the most likely contingency is the trip of one chiller, and that after it the remaining capacity, estimated from today's measured temperatures and flows, is $\dot{Q}_{\mathrm{avail},c} \sim \mathcal{N}(6.6, 0.25^2)$ MW, while the requirement, estimated from today's IT load forecast, is $\dot{Q}_{\mathrm{req},c} \sim \mathcal{N}(6.05, 0.10^2)$ MW. If the two are independent, the margin is normal with mean 0.55 MW and standard deviation $\sqrt{0.25^2 + 0.10^2} = 0.269$ MW, and

$$\Pr(M_c < 0) = \Phi\left(-\frac{0.55}{0.269}\right) = \Phi(-2.04) = 0.021,$$

where Φ is the standard normal cumulative distribution function. The same plant on a cool night, with a larger mean margin, has a probability orders of magnitude smaller. The quantity is a probability conditioned on present evidence, and it is the operational form of the adequacy question $\Pr(\dot{Q}_{\mathrm{avail}} < \dot{Q}_{\mathrm{req}})$.

12.6.2 Expected risk

A facility faces several contingencies, not one. Let c_1 be a chiller failure, c_2 a pump failure, c_3 the loss of a power path, and so on, each with a probability $\Pr(c_i)$ over the period of interest and a consequence ℓ_i, measured as expected downtime, lost compute, financial loss or the severity of a thermal violation. The expected risk is the probability-weighted sum of the consequences,

$$\mathcal{R} = \sum_i \Pr(c_i)\, \ell_i. \tag{12.15}$$

The consequence of each contingency is itself an outcome of the analysis of Section 12.4: it depends on the margin after the contingency, the time to limit and the success of the response.

Probability and consequence are separate. Risk is not probability. A contingency with $\Pr = 0.1$ and $\ell = 1$ contributes 0.1 to $\mathcal{R}$, and so does one with $\Pr = 0.001$ and $\ell = 100$: a hundred times rarer, a hundred times worse, the same contribution. A contingency with $\Pr = 0.001$ and $\ell = 1000$ contributes 1, ten times the first, although it is a hundred times less likely. Ranking contingencies by probability alone, as a list of the most frequent alarms does, puts rare and severe events such as common-cause failures at the bottom of a list on which they belong near the top.

12.6.3 Risk attribution

Equation (12.15) is a sum, so it can be split by contribution: which component, or which condition, contributes most to the expected risk. Table 12.5 does this for five illustrative sources: an aging UPS battery that may fail to carry the load during a transfer, the degradation of chiller 3 that erodes margin on hot days, an overloaded power path whose breaker may trip, a

fuel system common to all generators, and an isolation valve whose state is uncertain. Each row gives the annual probability, the consequence in hours of critical-load interruption, their product and its share of the total. The most likely source, chiller 3, contributes a sixth of the risk; the least likely, the common fuel system, contributes a third, because its consequence is a long outage of the whole facility.

Table 12.5: Risk attribution for five illustrative sources. $\Pr(c_i)$ is an annual probability, ℓ_i the consequence in hours of critical-load interruption, $\mathcal{R} = \sum_i \Pr(c_i)\ell_i = 0.296\,\mathrm{h}$ per year.

Source c_i	$\Pr(c_i)$	ℓ_i (h)	$\Pr(c_i)\,\ell_i$ (h)	Share
Common fuel system contaminated	0.002	48	0.096	32 %
Aging UPS battery fails at transfer	0.02	4	0.080	27 %
Chiller 3 degraded, margin lost on a hot day	0.10	0.5	0.050	17 %
Isolation valve state uncertain, stuck closed	0.05	1	0.050	17 %
Overloaded power path, breaker trip	0.01	2	0.020	7 %
Total			0.296	100 %

Attribution guides maintenance priorities. Replacing the UPS battery removes a quarter of the risk; testing the fuel removes most of a third; overhauling chiller 3, the item that generates the most alarms, removes a sixth. Verifying the valve costs little and removes a sixth, which makes it the cheapest item on the list per unit of risk. The attribution is only as good as the probabilities in it, and the next subsection shows that some of them are uncertain in a way that measurements can reduce.

12.6.4 Sensor and topology uncertainty

The inputs of a reliability calculation are not known exactly. A standby pump's status point may report AVAILABLE when the pump would not start: its breaker is racked out, its isolation valve is closed, or the status contact is wired to the wrong device. The status is measured; whether the pump would deliver flow on demand is inferred. Suppose that records of such status points show them to be wrong 5 % of the time (illustrative). Then, under the contingency of a duty-pump trip, the available cooling is not the single value 8 MW but a distribution,

$$\dot{Q}_{\mathrm{avail}} \mid y = \begin{cases} 8\,\mathrm{MW} & \text{with probability } 0.95, \\ 0 & \text{with probability } 0.05, \end{cases} \tag{12.16}$$

where y is the evidence, here the status reading and the history of such readings. The probability of a shortfall after the trip is 0.05, not zero. A capacity computed from status points is a calculated quantity resting on measurements that may be wrong, and the reliability built on it is itself uncertain. A test start of the standby pump would replace the 0.95 by a value near one, which is the sense in which a measurement reduces risk.

Hidden topology errors. The same applies to the topology. The facility's records say that rack 5 has independent A and B feeds. After a wiring change, both of its PDUs are in fact fed through one upstream breaker. The rack is $2N$ in the database and not physically. With feeds of reliability 0.99 each and a breaker of reliability 0.999, the rack's supply has reliability

$1 - 0.01^2 = 0.9999$ as recorded and $0.999 \times 0.9999 = 0.9989$ as built, a failure probability eleven times higher than the records imply, and no reliability calculation from the records can see it. The topology is an uncertain object: which rack is on which power path, which chiller on which switchboard, which valve open. Validating it, by comparing the currents measured on each path with the loads the recorded topology assigns to it, or by a controlled transfer test, is a reliability problem.

12.6.5 Graceful degradation and response policy

A facility that cannot carry its full load after a failure need not lose all of it. Graceful degradation is the deliberate reduction of service in steps, 100 % of compute to 80 % to 60 %, chosen to avoid a complete outage. The means are those of the recovery actions: shed noncritical loads, slow training jobs or other deferrable work, migrate workloads to another hall or site, and reduce redundancy temporarily, for example by running the remaining units without a spare until the failed one is repaired. In the example of Section 12.4.4, a cut from 100 % to 80 % of a 10 MW IT load brings the heat release to 8 MW, equal to the remaining capacity, and the facility runs on at reduced service instead of stopping.

Graceful degradation makes reliability partly a control problem. The consequence of a failure is a function of the failure, of the state of the facility when it occurs, of the controls applied and of how fast they are applied,

$$\ell = f(\text{failure},\ x,\ u,\ t_{\text{response}}), \tag{12.17}$$

with x the state and u the control actions. A better response policy therefore reduces risk without adding hardware. For the chiller contingency of Figure 12.7, suppose a manual response misses the fifteen-minute window with probability 0.2 and an automatic load-reduction sequence misses it with probability 0.01 (illustrative). If the window is missed the consequence is a thermal shutdown; otherwise it is a period at reduced service. The automatic sequence reduces the probability of the severe consequence, and with it that contingency's contribution to $\mathcal{R}$, by a factor of twenty, with no change to the plant.

12.6.6 The supervisory question

After a failure, the question an operator faces is not the redundancy level. It is: given what has failed and what the measurements show, what is the safest feasible operating state now? That question has concrete parts: which racks stay online, how much IT load can be supported, which cooling setpoints to use, and how long the facility can remain in the degraded mode before a second failure would cause an outage. Stated as a decision, it is to choose the actions u that keep the most service while holding the risk below a tolerance ϵ,

$$\begin{aligned} &\max_{u}\ P_{\text{IT}}(u) \\ &\text{subject to}\quad \Pr\!\left(\dot{Q}_{\text{avail},c} < \dot{Q}_{\text{req},c}(u) \mid y\right) \le \epsilon \quad \text{for every credible next contingency } c, \end{aligned} \tag{12.18}$$

together with the temperature limits over the time it takes to restore the failed equipment. Every element of the chapter enters it: the topology, the state and available capacity of each unit, the shared causes, the load, the weather and the response capability. Evaluating the probability is the subject of Chapter 17, testing candidate actions before applying them that of Chapter 18, and choosing among them that of Chapter 20.

> **Principle 12.5.** *$N{+}1$ or $2N$ describes the architecture; the risk at a given moment depends on topology, equipment state, available capacity, common causes, load, weather and response capability. The operational question is which failures the facility, as it exists now, can actually tolerate, and the answer is a probability conditioned on the present evidence.*

12.7 What to retain

Definition 12.1 (Operational reliability)**.** A redundancy level counts spare units; reliability is the probability that, for every credible contingency, an adequate response completes before available capacity falls short of the requirement for long enough to reach a limit. It depends on topology, shared causes, the available capacity of each unit under present conditions, the load and the response, and it is stated as a probability conditioned on present measurements, not as a property of the design.

Table 12.6 lists the chapter's objects and the subsection that defines each.

Table 12.6: The objects of redundancy and reliability analysis and where each is defined.

Object	Expression	Role	Defined in
Redundancy levels	N, $N{+}1$, $N{+}2$, $2N$; eq. (12.1)	spare units by design	12.1.1–12.1.3
Single point of failure	cut vertex of the dependency graph	shared element on every path	12.2.1, 12.2.2
Series reliability	$R_{\text{series}} = \prod_i R_i$	all components required	12.2.3
Parallel reliability	$1 - \prod_i (1 - R_i)$	any one suffices	12.2.3
k of n	equation (12.4)	$N{+}1$ and $N{+}2$ plants	12.2.3
Common cause	$q + (1 - q)p'^2$	dependence of redundant paths	12.2.4
Concurrent maintainability	any one unit out, load carried	planned work without outage	12.3.1
Fault tolerance	automatic continuation after failure	unplanned failure without outage	12.3.2
Availability A	$\text{MTBF}/(\text{MTBF} + \text{MTTR})$	fraction of time up	12.3.3
Downtime per year	$(1 - A) \times 8760\,\text{h}$	the nines	12.3.4
Reliability over t	$R(t) = e^{-t/\text{MTBF}}$	no failure in an interval	12.3.5
Available capacity	$\dot{Q}_{\text{avail}} = \sum_i \dot{Q}_{\text{avail},i}$	capacity states, derating	12.3.6
Capacity margin	$M_c = \dot{Q}_{\text{avail},c} - \dot{Q}_{\text{req},c}$	contingency test	12.4.1
Time to limit	$(T_{\max} - T)/(\mathrm{d}T/\mathrm{d}t)$	window for recovery	12.4.4
Event tree	products along branches	sequences after an initiating event	12.5.2
Fault tree	AND and OR gates	causes of a top event	12.5.3
Expected risk	$\mathcal{R} = \sum_i \Pr(c_i)\,\ell_i$	probability times consequence	12.6.2
Uncertain capacity	$p(\dot{Q}_{\text{avail}} \mid y)$	sensor and topology uncertainty	12.6.4
Supervisory problem	equation (12.18)	safest feasible state now	12.6.6

12.8 Exercises

Exercise 12.1. A data hall requires $\dot{Q}_{\text{req}} = 9\,\text{MW}$ of cooling from chillers rated 2.5 MW. (a) Compute N and the installed capacity of an $N{+}1$ plant. (b) Compute the margin M_c after the loss of one chiller. (c) On a hot day each chiller is derated to 2.1 MW; compute the capacity with all units in service and the margin after one failure. (d) What is the lowest per-chiller capacity at which the $N{+}1$ plant still survives one failure, and what derating, in percent of the rating, does that allow?

Exercise 12.2. One electrical path consists of a utility feed, a UPS and a PDU in series, with reliabilities 0.995, 0.99 and 0.998 over a year. (a) Compute the reliability of one path. (b) Compute the reliability of two such paths in parallel, assuming independence. (c) Both paths are in fact fed through one common switchboard of reliability 0.999; compute the reliability of the arrangement and the ratio of its failure probability to that of (b). (d) Instead of the switchboard, the two UPS rooms share one cooling system that fails with probability $q = 0.002$ and then takes both paths down; holding each path's failure probability at the value from (a), use equation (12.6) with that value in place of 0.01 to compute the probability that both paths fail.

Exercise 12.3. A UPS has MTBF = 20 000 h and MTTR = 8 h. (a) Compute its availability, its downtime per year, and its number of nines. (b) Compute the availability of two such units in parallel with independent failures and repairs, and the downtime per year in minutes. (c) A second design has MTBF = 2000 h and MTTR = 0.8 h. Show that its availability equals that of (a), and compute the reliability of each over 1000 h with a constant failure rate. Which is preferable for a workload that loses an hour of computation at each interruption, and by how much per year?

Exercise 12.4. A hall has $\dot{Q}_{\mathrm{req}} = 12\,\mathrm{MW}$ and an effective thermal capacitance of 400 MJ/K. A chiller trip leaves $\dot{Q}_{\mathrm{avail}} = 9\,\mathrm{MW}$; the rack inlet is at 22°C with a limit of 27°C. (a) Compute the rate of rise in °C per minute and the time to limit. (b) A standby chiller takes 8 min to come on line and starts with probability 0.9. If it fails, the operators can cut IT load; what fraction of the load must be cut to stop the rise, assuming heat release equals IT power? (c) Assign a consequence of 0.25 h of interruption if the standby starts, 1 h if it fails and the load cut is made in time, and 8 h if the cut is too late; the cut is too late with probability 0.3 under manual response and 0.02 under an automatic sequence. With a chiller-trip probability of 0.4 per year, compute the contribution of this contingency to $\mathcal{R}$ under each response, and the reduction factor.

12.9 Connection to the TERRA framework

Redundancy and reliability are statements about the graph of the facility under the removal of its parts, and they map onto the TERRA framework as follows.

- **Redundant units are parallel components with capacity bounds.** Each chiller, pump, UPS and PDU is a component on an edge of the electrical, water or thermal network, with an available capacity $\dot{Q}_{\mathrm{avail},i}$ as its bound (equation (12.9)). A redundancy level is not stored as a label; it emerges from the bounds and the topology, so the degraded $N{+}1$ of Figure 12.6 appears as soon as the bounds are updated for the weather.

- **A contingency is a modified graph, and the margin is a solved balance.** Removing a component and solving the three coupled conservation networks again gives M_c of equation (12.11) for every row of Table 12.4, including the switchboard row in which an electrical removal produces a thermal shortfall. Single points of failure are the nodes whose removal disconnects the racks from every source, as in Figure 12.2.

- **Shared causes are shared parent nodes.** The cooling of the UPS rooms in Figure 12.4 is a thermal node on which two electrical components depend; representing it as such, rather than assigning independent failure rates, is what makes equation (12.6) come out of the model instead of being assumed.

- **Time to limit is forward simulation of a thermal node.** After a contingency, the room heat balance of equation (12.13) is simulated forward with the recovery actions as

control inputs, which produces trajectories such as those of Figure 12.7 and the time window that the event tree of Figure 12.8 assumes.

- **Capacity, status and topology are inferred, and risk is a posterior probability.** The available capacity of each unit, the readiness of a standby pump (equation (12.16)) and the feed of rack 5 are inferred from measurements, not read from nameplates or records; the risk of Section 12.6.1 is the posterior probability $\Pr(\dot{Q}_{\mathrm{avail},c} < \dot{Q}_{\mathrm{req},c} \mid y)$, and the measurement that most reduces it, a pump test start or a transfer test, is chosen for the information it adds.

In summary, redundancy is a count of spare units and reliability is a probability: series dependence, shared causes, maintenance, degraded capacity, the time a shortfall takes to reach a limit, and the speed of the response all separate the two, and the operational answer is the probability, conditioned on present measurements, that the facility as it now exists tolerates its credible contingencies. The next chapter, Chapter 13, turns from what failures cost to how they are recognized: the failure modes of data center equipment and the symptoms by which each is diagnosed.

Chapter 13

Failure Modes and Diagnostics

13.1 Symptoms, causes and hypotheses

The earlier chapters described how a data center behaves when it works: how power reaches the IT equipment, how the heat it produces is carried by air and water to the outdoors, what the sensors report of that chain and how the controls act on it. Chapter 12 treated failure as an event with a probability. This chapter turns to what happens in the telemetry when something does go wrong, and to the reasoning that leads from an abnormal reading back to its cause. The question is

given abnormal telemetry, how are competing physical causes of the same symptom distinguished from one another?

The difficulty is that a symptom is not a cause. A rising rack inlet temperature may come from a loss of airflow, from hot air recirculating to the inlet, from a higher IT load, from warm chilled water, or from a temperature sensor that reads high. Each cause calls for a different action, and the reading itself does not say which one is present. The chapter first catalogues the failure modes of a data center, air side, water side, sensors, topology and electrical equipment, each on the same template; it then develops the residual-based and probabilistic reasoning that ranks the candidate causes, and ends with the step from a ranked diagnosis to an action.

Jargon: symptom, fault, failure mode, hypothesis, signature, residual, prior, likelihood, posterior, fault detection and isolation, anomaly detection, latent state

A *symptom* is an observed departure from normal behavior, such as a temperature above its usual value. A *fault* is the physical cause of the departure: a degraded fan, a stuck valve, a biased sensor. A *failure mode* is a named way in which a component or subsystem can fail, together with its mechanism. A *hypothesis* H_i is a candidate explanation of a set of observations; H_0 denotes normal operation. The *signature* of a fault is the pattern of changes it produces across all measured variables, in direction, size, place and time. A *residual* r is the difference between a measured value and the value a model predicts. In probabilistic reasoning, the *prior* $\Pr(H_i)$ is the probability of a hypothesis before the current measurements are considered, the *likelihood* $\Pr(y \mid H_i)$ is the probability of the measurements y if the hypothesis is true, and the *posterior* $\Pr(H_i \mid y)$ is the probability of the hypothesis after the measurements are considered. *Fault detection and isolation* (FDI) is the engineering discipline of deciding that a fault is present and locating it. *Anomaly detection* flags behavior that is statistically unusual without explaining it. A *latent state* is a physical state that normal telemetry does not reveal, such as the true usable capacity of a battery that reports itself fully charged.

Key idea

Diagnosis does not ask which value crossed a threshold; it asks which physical explanation best accounts for all the available evidence at once. Each candidate cause predicts its own pattern across temperatures, powers, flows, pressures, places and times, and the measured pattern is compared against every prediction, weighted by how often each cause occurs and by how much it would cost to be wrong.

13.1.1 Symptom and cause

Take a rack whose inlet temperature has risen from its usual 22°C to 27°C. The rise is *measured*: a sensor reports it. What produced it is not measured and must be *inferred*, and at least five physical explanations are possible. The IT load may have increased, raising the rack's exhaust temperature and, through recirculation, its inlet. The rack's airflow may have fallen. Hot exhaust air may be reaching the inlet by a new path. The computer room air handler (CRAH) may be delivering warmer or less air. The chilled water feeding that CRAH may be warmer than usual. Or the physical temperature may be unchanged and the sensor may simply read 5 K high.

A threshold alarm treats these explanations identically: it compares one number with a limit and fires. The comparison is useful for detection, but it carries no information about the cause, and the response to each cause is different. A degraded fan is replaced, a blocked filter cleaned, a recirculation path closed with blanking panels, a chilled-water problem escalated to the plant, and a biased sensor recalibrated while the cooling is left alone. Diagnosis is the reasoning that separates these cases, and its raw material is not the symptom but everything else the telemetry reports at the same time.

13.1.2 Competing hypotheses

The explanations of the inlet rise are written as hypotheses:

- H_1: the IT load of the rack increased;
- H_2: the rack airflow is reduced;
- H_3: hot exhaust air is recirculating to the inlet;
- H_4: the CRAH serving the rack has degraded;
- H_5: the chilled water supplied to that CRAH has a problem;
- H_6: the inlet temperature sensor is biased.

Normal operation, H_0, is the hypothesis that the readings are within the range the normal model predicts. The diagnostic task is not to find a hypothesis that explains the inlet temperature, since all six do, but to find the one that best explains the *entire* set of measurements: rack power, fan speeds, pressures, supply air and water temperatures, neighboring racks, and the timing of each change.

Figure 13.1 draws the structure of the problem. The box on the left is the single symptom. The six boxes in the middle are the hypotheses, one per arrow. The column on the right lists, for each hypothesis, the evidence it predicts beyond the symptom, which is what distinguishes it from the others. Read from right to left, the figure is the diagnostic question: which of the right-hand patterns is present in the data. Every right-hand entry involves a variable other than the inlet temperature, so a diagnosis that looks at the inlet temperature alone cannot choose among the six.

13.1.3 Signatures

Each hypothesis predicts a *signature*: a pattern of changes across the measured variables. Under a high IT load (H_1), rack power P_{rack} rises, the exhaust temperature T_{out} rises with it, and

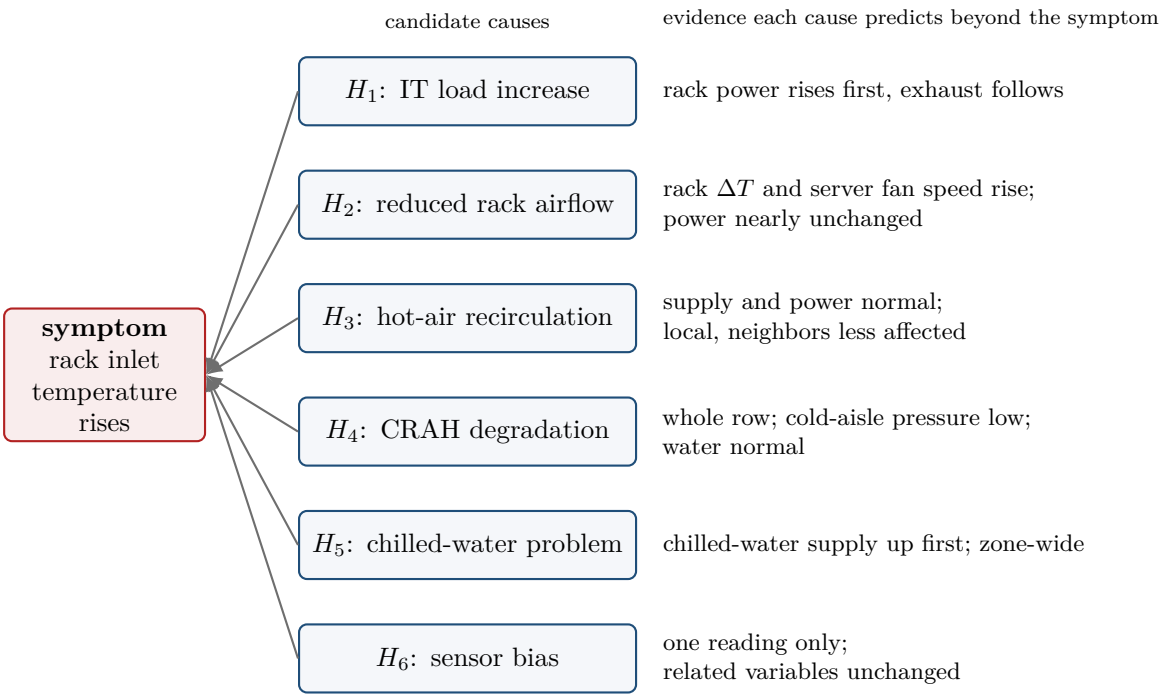


Figure 13.1: One symptom, six competing hypotheses. Arrows run from each candidate cause to the common symptom; the right column lists the additional evidence each cause predicts, which is what separates it from the others. The inlet temperature alone cannot choose among them.

the inlet temperature T_{in} may rise after a delay as some of the hotter exhaust finds its way back to the inlet. Under low airflow (H_2), the rack mass flow $\dot{m}_{\text{air}}$ falls, the rack temperature rise $\Delta T_{\text{rack}} = T_{\text{out}} - T_{\text{in}}$ grows, and the server fans speed up as their controllers react to warmer components, while P_{rack} hardly moves. Under sensor bias (H_6), the biased reading rises while every physically related variable, the neighbors' readings, rack power, fan speeds, supply temperatures, stays where it was.

Diagnosis is therefore pattern matching: the measured pattern is compared with the pattern each hypothesis predicts, and the hypotheses whose predictions disagree with the data lose credibility. Two requirements follow. The first is a model for each hypothesis that predicts its signature, which is what the physics of the earlier chapters supplies. The second is measurements on which the signatures differ: if two hypotheses predict the same values for every measured variable, no amount of reasoning separates them. The catalogue of the next three sections states each failure mode on one template, its *mechanism*, the *governing relation* that turns the mechanism into numbers, the *signature*, and *what distinguishes it* from the failure modes it resembles.

> **Principle 13.1.** *A symptom is not a fault. Many physical causes produce the same symptom; they differ in the pattern they produce across all the other measured variables, and that pattern, not the symptom, is what a diagnosis reads.*

13.2 Air-side failure modes

The air side moves heat from the servers to the CRAH coils. Its failure modes change either how much air reaches the IT equipment or what temperature that air has; this section treats four of them, from the smallest footprint to the largest.

13.2.1 Server or rack fan degradation

Mechanism. A server fan loses speed or pressure through bearing wear, a failed rotor in a multi-rotor assembly, dust on the blades, or a fault in its speed controller. The server then moves less air, $\dot{m}_{\text{server}}$ falls, while the heat the server produces is set by its computational load and stays nearly constant.

Governing relation. The heat carried by the air is $\dot{Q} = \dot{m} c_p \Delta T$ (Chapter 5). At constant heat, the temperature rise across the server is inversely proportional to its airflow,

$$\Delta T_{\text{server}} = \frac{P_{\text{server}}}{\dot{m}_{\text{server}}\, c_{p,\text{air}}}. \tag{13.1}$$

For an illustrative server of $P_{\text{server}} = 500\,\text{W}$ with $\dot{m}_{\text{server}} = 0.040\,\text{kg/s}$ and $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$, the rise is $0.5/(0.040 \times 1.0) = 12.5\,\text{K}$. If the fans lose 30 % of their flow, $\dot{m}_{\text{server}} = 0.028\,\text{kg/s}$ and the rise becomes $0.5/0.028 = 17.9\,\text{K}$. With the inlet at 22°C the exhaust moves from 34.5 to 39.9°C. The processor temperature T_{CPU} rises by more than the exhaust: it sits above the local air temperature by its heat times a thermal resistance, and that resistance grows as the air velocity over the heat sink falls.

Signature. Server exhaust and processor temperatures rise; the remaining fans of the same server speed up as the server's controller reacts; server power rises slightly with the extra fan power; rack power changes only slightly; the rack inlet temperature stays normal; neighboring servers are unaffected.

What distinguishes it. The footprint is one server, the inlet is normal and the facility cooling is not involved: this is a local server problem, not a facility cooling problem. When the server's management controller reports individual fan speeds, a failed or slow fan is *measured* directly; when only the server's exhaust and component temperatures are available, the fan loss is *inferred* from a temperature rise at a normal inlet and a constant power.

13.2.2 Blocked airflow

Mechanism. Cable bundles packed behind the servers, a clogged filter or another obstruction in the rack's air path raise the flow resistance of the path. Missing blanking panels act on the same flow path but mainly by opening a route for exhaust air back to the front, which is the recirculation of Section 13.2.3.

Governing relation. The pressure drop of a flow path is $\Delta p = K\dot{V}^2$ (Chapter 6), and the flow settles where the pressure the fans can produce equals the pressure the path demands, the intersection of the fan curve and the system curve. A blockage raises K and moves the intersection to a lower flow. Take the rack's server fans together as one illustrative equivalent fan with the curve $\Delta p = 400 - 100\,\dot{V}^2$ (pascals, with $\dot{V}$ in m^3/s) at nominal speed, and a clean path with $K = 150\,\text{Pa}/(\text{m}^3/\text{s})^2$. The operating point solves $150\,\dot{V}^2 = 400 - 100\,\dot{V}^2$, so $\dot{V}^2 = 1.6$, $\dot{V} = 1.265\,\text{m}^3/\text{s}$ and $\Delta p = 240\,\text{Pa}$. A blockage that doubles the resistance to $K = 300$ gives $400\,\dot{V}^2 = 400$, $\dot{V} = 1.0\,\text{m}^3/\text{s}$ at $\Delta p = 300\,\text{Pa}$: 21 % less flow at the same fan speed.

With an air density of $1.2\,\text{kg/m}^3$, the mass flow falls from 1.52 to 1.20 kg/s, and for an illustrative rack heat of 20 kW the rack temperature rise grows from $20/(1.52 \times 1.0) = 13.2\,\text{K}$ to $20/1.20 = 16.7\,\text{K}$. The server fans respond by speeding up. By the affinity laws (Chapter 6) the fan pressure scales with the square of the speed ratio s and the free-delivery flow with s, so the fan curve becomes $\Delta p = 400\,s^2 - 100\,\dot{V}^2$. Restoring $1.265\,\text{m}^3/\text{s}$ through the blocked path requires $300 \times 1.6 = 480\,\text{Pa}$, hence $400\,s^2 = 480 + 160$, $s^2 = 1.6$ and $s = 1.265$: 26.5 % more

speed. The fan power, which is pressure times flow divided by efficiency, rises by the factor $480/240 = 2.0$. If the fans drew 0.76 kW of the rack's 20 kW before the blockage, they draw about 1.5 kW after it, a change of 0.76 kW, or 4 % of the rack power.

Figure 13.2 draws these curves. The horizontal axis is the rack airflow in cubic meters per second; the vertical axis is the pressure difference in pascals. The two falling curves are the fan curve at nominal speed (solid) and at the raised speed (dashed); the two rising curves are the system curve of the clean path and of the blocked path. Point A is the design operating point, point B the blocked path at nominal speed, and point C the blocked path after the fans have sped up to restore the flow. The move from A to B is the airflow loss; the move from B to C is the fan's compensation, paid for in pressure, speed and fan power.

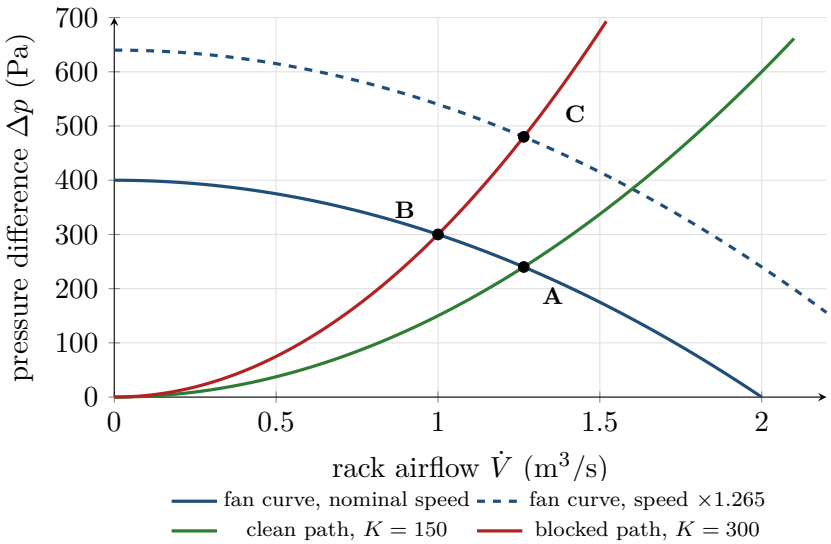


Figure 13.2: Blocked airflow as a shift of the operating point, illustrative. Fan curves $\Delta p = 400\,s^2 - 100\,\dot{V}^2$ with $s = 1$ and $s = 1.265$; system curves $\Delta p = K\dot{V}^2$ with $K = 150$ (clean) and $K = 300\,\mathrm{Pa/(m^3/s)^2}$ (blocked). A ($1.27\,\mathrm{m^3/s}$, 240 Pa) is the design point; a blockage alone moves the rack to B ($1.00\,\mathrm{m^3/s}$, 300 Pa), 21 % less flow; the fans recover the flow at C ($1.27\,\mathrm{m^3/s}$, 480 Pa) by running 26.5 % faster at twice the fan power.

Signature. The rack temperature rise ΔT_{rack} grows, the server fan speeds N_{fan} rise, and the pressure drop across the filter, $\Delta p_{\mathrm{filter}}$, rises when the blockage is a clogged filter and that pressure is measured. Rack power is almost unchanged: the extra fan power of the example is under 1 kW.

What distinguishes it. Against a high IT load (H_1), the rack power: a load increase that warms the rack by the same amount adds kilowatts of IT power, while a blockage adds only fan power. Against a weak fan (Section 13.2.1), the pressure: a blockage raises the pressure drop across the obstruction while the fans run faster, whereas a weak fan delivers less flow at a lower pressure, so the pressure drop across the same path falls.

13.2.3 Hot-air recirculation

Mechanism. Exhaust air returns to rack inlets over the tops of racks, around the ends of rows, through empty rack positions and missing blanking panels, or because the supply delivers less air than the racks draw. The servers' airflow is normal; what changes is the temperature of the air they take in.

Governing relation. With a recirculation fraction α, the share of the inlet air that is exhaust, the inlet temperature is the mixture $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{hot}}$ (Chapter 6). For an illustrative supply at $T_{\text{supply}} = 20°\text{C}$ and exhaust at $T_{\text{hot}} = 36°\text{C}$, a fraction of $\alpha = 0.05$ gives $T_{\text{in}} = 0.95\times20+0.05\times36 = 20.8°\text{C}$, and a fraction of $\alpha = 0.30$ gives $0.70\times20+0.30\times36 = 24.8°\text{C}$, a rise of 4 K. The sensitivity is $\partial T_{\text{in}}/\partial\alpha = T_{\text{hot}} - T_{\text{supply}} = 16\,\text{K}$, or 0.16 K for each percentage point of recirculation.

Signature. The inlet temperature rises; the CRAH supply temperature is normal; the rack power is unchanged; the local hot aisle is elevated; the affected rack, or the rack at the end of a row, is worst and its neighbors are progressively less affected. The rise is often largest at the top of the rack, where exhaust air spills over.

What distinguishes it. A normal supply temperature rules out the CRAH and the plant; an unchanged rack power rules out the load; an unchanged rack temperature rise rules out a loss of rack airflow, since recirculation lifts the inlet and the exhaust together. What remains is spatial: the problem is the path between racks, a topological fact about the room, and its signature is a spatial pattern rather than a change in any one device. The fraction α itself is never measured; it is *inferred* from the inlet, supply and exhaust temperatures through the mixing relation.

13.2.4 CRAH fan degradation

Mechanism. A CRAH moves less air because of belt slip, a motor or variable-frequency-drive fault, bearing wear, or the failure of one module of a fan array.

Governing relation. The effect travels as a chain,

$$\dot{V}_{\text{CRAH}} \text{ falls} \rightarrow \text{cold-aisle pressure falls} \rightarrow \text{flow through the tiles and racks falls} \rightarrow T_{\text{in}} \text{ rises.}$$

When the supply flow drops below what the racks draw, the racks' fans make up the difference with room air, and in the worst case all of it is exhaust, so the recirculation fraction is at least $\alpha \approx 1 - \dot{V}_{\text{supply}}/\dot{V}_{\text{racks}}$. In an illustrative row whose racks draw $9\,\text{m}^3/\text{s}$ from a CRAH supplying $10\,\text{m}^3/\text{s}$, a 20 % loss of CRAH flow leaves $8\,\text{m}^3/\text{s}$, so $\alpha \approx 1 - 8/9 = 0.11$ on average and $T_{\text{in}} \approx 20 + 0.11 \times 16 = 21.8°\text{C}$, with the racks at the ends of the row, farthest from the supply, well above the average.

Signature. Inlet temperatures rise across the whole row the CRAH serves; the cold-aisle pressure is low; the CRAH's fan feedback is abnormal (lower power, or a high speed command with a low flow); the chilled-water supply and return temperatures are fairly normal, and the CRAH supply air may even be slightly colder, since less air passes the same coil.

What distinguishes it. It separates the air side from the water side: the cooling is available in the coil but the air does not carry it to the racks. Against recirculation at one rack, the footprint is the row and the cold-aisle pressure is low; against a chilled-water problem, the water temperatures and the CRAH supply air temperature are normal.

13.3 Water-side and plant failure modes

The water side moves heat from the CRAH coils to the chillers and from the chillers to the cooling towers. Its failure modes have wider footprints than those of the air side, because each element upstream serves many elements downstream; this section treats four of them and then the spatial pattern that locates them.

13.3.1 Stuck chilled-water valve

Mechanism. The control valve that sets the chilled-water flow through a CRAH coil sticks: a seized stem, a failed actuator, a broken linkage. The controller commands one position and the valve holds another.

Governing relation. The heat the coil removes is limited by the water that passes it, $\dot{Q}_{\text{coil}} = \dot{m}_w c_{p,w}(T_{\text{CHW,r}} - T_{\text{CHW,s}})$, and the same heat leaves the air, $\dot{Q}_{\text{coil}} = \dot{m}_{\text{air}} c_{p,\text{air}}(T_{\text{return}} - T_{\text{supply}})$ (Chapter 7). Take an illustrative coil that passes $\dot{m}_w = 8\,\text{kg/s}$ at a commanded opening of 80 % and warms it by 6 K: $\dot{Q}_{\text{coil}} = 8 \times 4.18 \times 6 \approx 200\,\text{kW}$. With $\dot{m}_{\text{air}} = 14\,\text{kg/s}$ of air returning at 32°C the supply air is $32 - 200/(14 \times 1.0) \approx 17.7$°C. If the valve is stuck at 20 % and the flow is proportional to the opening, only 2 kg/s passes. The water can warm more, but not without limit, since it cannot leave the coil warmer than the air entering it; with an illustrative limit of 12 K the coil delivers at most $2 \times 4.18 \times 12 \approx 100\,\text{kW}$, and the supply air at the same return temperature is $32 - 100/14 \approx 24.9$°C, about 7 K warmer.

Signature. The supply air of that one CRAH is warm. Its controller, seeing the supply above setpoint, drives the command toward 100 %, so the command is high and the temperature is high at the same time, the mark of an actuator that does not do what it is told. The water temperature rise across the coil is larger than usual, and the branch flow, where it is metered, is low. Other CRAHs on the same loop are normal.

What distinguishes it. The commanded state is not the physical state (Chapter 9). A valve position feedback, if it reports the command rather than the stem, does not reveal the fault; the heat balance across the coil does. Against a flow-sensor bias (Section 13.4.2), the air: a low flow reading with a normal supply air temperature points to the sensor, a low flow reading with a warm supply air points to the valve.

> **Principle 13.2.** *The commanded state of an actuator is not its physical state. A command that rises while its effect does not appear is itself evidence of an actuator fault.*

13.3.2 Pump degradation

Mechanism. A chilled-water pump delivers less pressure or flow because of impeller wear, a drive fault, cavitation, or the loss of one pump in a parallel set.

Governing relation. The loop obeys the same quadratic as the air path, $\Delta p_{\text{loop}} = K_{\text{loop}} \dot{V}^2$, so at fixed valve positions the flow scales with the square root of the pump pressure. A pump that loses 20 % of its pressure reduces the loop flow by a factor $\sqrt{0.8} = 0.894$, about 11 %. Every coil on the loop then carries less water; at an unchanged heat load the water temperature rise across each coil grows by the inverse factor, from 6 K to $6/0.894 = 6.7$ K, and the return water is about 0.7 K warmer. Coils that are hydraulically far from the pump lose more than the average.

Signature. The consequences are widespread: several CRAHs lose flow at once, their valve commands rise together, the chilled-water return temperature shifts upward, the loop pressure is low, and if the valves run out of travel the zone served by the loop warms. The pump's speed and power readings may be abnormal.

What distinguishes it. The spatial pattern. A fault seen at every CRAH on one loop, and at no CRAH on another loop, points to the element the affected CRAHs share, which is upstream of all of them. Against a stuck valve, the number of coils affected; against a chiller problem, the supply water temperature, which a pump fault leaves normal until the chiller itself is starved of flow.

13.3.3 Chiller degradation

Mechanism. Fouling of the evaporator or condenser tubes, compressor wear, a loss of refrigerant charge, or a condenser problem lowers the chiller's efficiency or its capacity.

Governing relation. The coefficient of performance is COP $= \dot{Q}_{\text{evap}}/W_{\text{comp}}$, and the condenser rejects $\dot{Q}_{\text{cond}} = \dot{Q}_{\text{evap}} + W_{\text{comp}}$ (Chapter 7). For an illustrative chiller removing $\dot{Q}_{\text{evap}} = 2\,\text{MW}$ at COP $= 5$, the compressor draws 400 kW. If degradation lowers the COP to 4.2, the same cooling costs $2000/4.2 = 476\,\text{kW}$, 19 % more. The capacity falls too: with an illustrative compressor limit of 500 kW the chiller can remove at most $5 \times 500 = 2.5\,\text{MW}$ before degradation and $4.2 \times 500 = 2.1\,\text{MW}$ after it. Under a load of 2.2 MW the degraded chiller can no longer hold its chilled-water supply setpoint, and $T_{\text{CHW,s}}$ rises.

Signature. The compressor power per unit of cooling rises (COP falls) at an unchanged condenser water temperature; or the available capacity $\dot{Q}_{\text{avail}}$ falls; and the chilled-water supply temperature rises only under high load, when the reduced capacity is reached. The COP is a *calculated* quantity: W_{comp} is metered and $\dot{Q}_{\text{evap}}$ is calculated from the chilled-water flow and temperature difference, so its accuracy depends on the flow meter.

What distinguishes it. The COP compared with the COP expected at the measured temperature lift. A chiller whose efficiency is low for its lift is degraded; a chiller whose efficiency is normal for an unusually large lift is suffering from the conditions around it, which is the next failure mode.

13.3.4 Cooling-tower degradation

Mechanism. Fouled or scaled fill, a failed tower fan, or poor water distribution reduces the tower's ability to cool the condenser water toward the outdoor wet-bulb temperature T_{wb}.

Governing relation. A tower's *approach* is the difference $T_{\text{CW,s}} - T_{\text{wb}}$ between the condenser-water supply temperature it delivers and the wet-bulb temperature. A degraded tower runs a larger approach, so $T_{\text{CW,s}}$ rises; the chiller's condensing temperature T_{cond} follows it; the lift $T_{\text{cond}} - T_{\text{evap}}$ grows; and by the fraction-of-Carnot model COP $\approx \eta_{\text{C}} T_{\text{evap}}/(T_{\text{cond}} - T_{\text{evap}})$ (Chapter 7) the COP falls. With the illustrative values $\eta_{\text{C}} = 0.5$, $T_{\text{evap}} = 278\,\text{K}$ and a normal lift of 27.8 K, COP $= 0.5 \times 278/27.8 = 5.0$. If the approach grows from 4 to 7 K, the lift grows by 3 K to 30.8 K and COP $= 139/30.8 = 4.51$. The compressor power for 2 MW of cooling rises from 400 to 443 kW, and the capacity at a 500 kW compressor limit falls from 2.5 to 2.26 MW.

Signature. Condenser-water supply temperature above what the wet-bulb temperature and the design approach predict; compressor power up; available capacity possibly down; at high load, chilled-water supply temperature up. An upstream heat-rejection fault propagates back through the chiller and the chilled water to rack risk.

What distinguishes it. The approach at the measured wet-bulb temperature. A large approach is a tower fault; a normal approach on a hot, humid day gives the same high condenser water and is weather, not a fault. Against chiller degradation, the chiller's COP is normal for its lift: the chiller is doing its job against a harder condition.

13.3.5 Footprint along the cooling chain

The failure modes of the last two sections sit at different points of one chain, and their position on the chain sets their spatial footprint. Heat travels from the servers through the rack and aisle, the CRAH fan and coil, the chilled-water loop and pumps, the chiller and the cooling tower to the outdoor air. A fault at a point of the chain affects every element downstream of it in the direction of cooling, which is everything that the faulty element serves.

Figure 13.3 draws the chain. The upper row is the air side and the lower row the water side and heat rejection; the arrows follow the heat from the servers to the outdoor air. Each box names an element and, on its second line, the footprint of a fault at that element: the set of racks whose temperatures it disturbs. The red labels are the failure modes located at each element. Reading the boxes in the direction of the arrows, the footprint widens from one server to the whole facility. Read in reverse, the figure is a locating rule: one rack points to a local fan, obstruction, recirculation path or sensor; a whole row points to a CRAH, a pump branch or a control zone; the whole facility points to the chiller plant, the utility supply, a global control, or the weather.

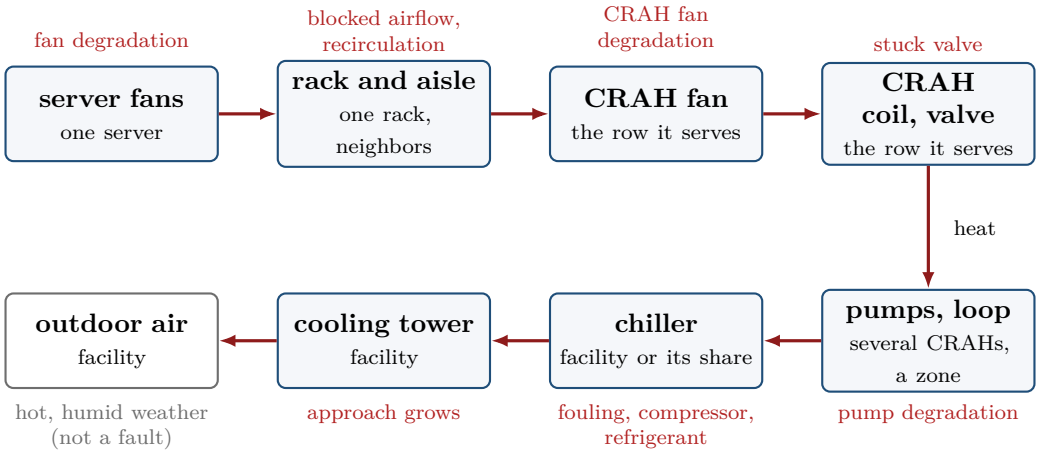


Figure 13.3: The cooling chain from the servers to the outdoor air, with the failure modes located at each element (red) and the footprint of a fault at that element (second line of each box). Arrows follow the heat. A fault upstream disturbs every rack downstream of it, so the size of the footprint locates the fault on the chain.

> **Principle 13.3.** *The spatial footprint of a fault locates it on the cooling chain: a fault at an element disturbs everything that element serves, so one rack, one row, one zone and the whole facility point to different elements.*

13.4 Sensor, topology and electrical failure modes

Not every fault is in the cooling equipment. The measurements themselves can be wrong, the model of which element serves which can be wrong, and the electrical system degrades in ways

that show up thermally or not at all; this section treats six such failure modes on the same template.

13.4.1 Bad temperature sensor

Mechanism. A temperature sensor acquires a bias b, through drift, a damaged element, a loose mounting that places it in a warmer air stream, or a configuration error in its scaling. With a true inlet temperature of 24°C and $b = +5\,\mathrm{K}$, it reads 29°C.

Governing relation. The reading is $y_T = T + b + \epsilon$ (Chapter 9), and no single reading separates T from b. What separates them is the physical relations the true temperature must satisfy together with other measured variables. Take the rack's energy balance $P_{\mathrm{rack}} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}} (T_{\mathrm{out}} - T_{\mathrm{in}})$ with an illustrative measured $P_{\mathrm{rack}} = 18\,\mathrm{kW}$ and $T_{\mathrm{out}} = 36°\mathrm{C}$. With the true inlet of 24°C the implied airflow is $18/(1.0 \times 12) = 1.5\,\mathrm{kg/s}$, the rack's normal value. With the biased reading of 29°C it would be $18/(1.0 \times 7) = 2.57\,\mathrm{kg/s}$, 71 % more air from fans whose speeds have not changed, which is physically impossible. Holding the airflow at its normal 1.5 kg/s instead and solving the balance for the inlet gives $36 - 12 = 24°\mathrm{C}$, and an *inferred* bias of $29 - 24 = 5\,\mathrm{K}$.

Signature. Neighboring inlet sensors read normally; rack power is unchanged; the airflow inferred from the other measurements is normal; the chilled-water and supply air temperatures are normal; the servers' internal inlet and component temperatures are normal. Only the one reading has moved.

What distinguishes it. The reading violates the expected physical relationships with every other measurement. This is *consistency-based diagnosis*: a sensor is suspected not because its value is high but because no physically possible state of the system produces that value together with all the other values.

13.4.2 Bad flow sensor

Mechanism. A water flow meter drifts, fouls, loses its signal or is affected by air in the line, and its reading falls while the true flow does not.

Governing relation. The coil's heat can be computed two ways: from the flow meter, $\dot{Q}_{\mathrm{flow}} = \dot{m}_w c_{p,w} (T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}})$, and from an independent estimate of the load the coil carries, $\dot{Q}_{\mathrm{expected}}$, such as the air-side balance or the IT power of the racks it serves. Take an illustrative coil whose meter used to read 10 kg/s at a water temperature rise of 6 K, which is $10 \times 4.18 \times 6 = 251\,\mathrm{kW}$, in agreement with the air side. The meter now reads 7 kg/s, while the temperature rise is still 6 K and the air side still says 251 kW. The flow-based heat is $7 \times 4.18 \times 6 = 176\,\mathrm{kW}$, a residual of 75 kW, 30 %. If the flow had really fallen to 7 kg/s at an unchanged load, the water temperature rise would have grown to $251/(7 \times 4.18) = 8.6\,\mathrm{K}$; it did not.

Signature. Measured flow down; water temperature rise unchanged; air-side and IT-side heat unchanged; supply air temperature normal.

What distinguishes it. A mismatch between $\dot{Q}_{\mathrm{expected}}$ and $\dot{Q}_{\mathrm{flow}}$ suggests a sensor error or a model error, and the temperatures decide which variable is at fault: a real flow loss raises the water temperature rise and warms the supply air, a flow-sensor fault changes neither.

13.4.3 Topology error

Mechanism. The monitoring system's model of the facility says that CRAH 3 serves row A. After a reconfiguration, new containment, moved tiles or a changed row layout, CRAH 3 mainly serves row B. No equipment has failed; the description of which element is connected to which, the topology G, is wrong.

Governing relation. The topology fixes the sensitivities the model expects: how much each row's temperature moves when each CRAH changes. In an illustrative test, raising the supply setpoint of CRAH 3 by 1 K is expected, under the recorded topology, to raise the inlets of row A by about 0.7 K and those of row B by about 0.1 K. The observed changes are 0.05 K in row A and 0.6 K in row B.

Signature. Changes at CRAH 3 do not affect the temperatures of row A as expected but do affect those of row B. Residuals appear when any element changes, and their pattern follows the true connections rather than any one device.

What distinguishes it. A degraded CRAH 3 would weaken its effect on row A without transferring that effect to row B; a topology error moves the effect from one row to another. The remedy is a corrected graph, not a repair, and until it is corrected every diagnosis that relies on the wrong graph will blame the wrong equipment. Topology is therefore itself a hypothesis to be tested, not a fixed input.

13.4.4 UPS degradation

Mechanism. An uninterruptible power supply (UPS) loses efficiency through aging capacitors, fan failures in its own cooling, or a degraded power stage.

Governing relation. The UPS loss is $P_{\text{in}} - P_{\text{out}} = P_{\text{out}}(1/\eta - 1)$ and all of it becomes heat in the electrical room, $\dot{Q}_{\text{UPS}}$. For an illustrative UPS delivering $P_{\text{out}} = 1\,\text{MW}$, an efficiency of $\eta = 0.96$ gives a loss of $1000(1/0.96 - 1) = 41.7\,\text{kW}$; a degraded $\eta = 0.93$ gives 75.3 kW, an extra 33.6 kW of heat that the electrical room's cooling must remove, at the cost of extra cooling power as well.

Signature. The difference between the input and output meters of the UPS rises at an unchanged output; the electrical room warms or its cooling works harder; the facility power rises with no change in IT power.

What distinguishes it. The fault is cross-domain: an electrical degradation appears first as a thermal load. When both UPS meters exist, the loss is *calculated* from them directly; when only one does, it is *inferred* from the electrical room's heat balance.

13.4.5 Battery degradation

Mechanism. The UPS battery loses usable capacity as its state of health SOH declines, while its state of charge SOC reports 100 %. The battery is full, but full of less than it used to hold. A battery designed for an illustrative ten minutes at full load (Chapter 4) that has fallen to SOH = 0.7 holds about seven minutes of energy, and at high discharge rates delivers less.

Signature and what distinguishes it. In normal float operation, nothing: the voltage is normal, the charge state is full and no current flows. The usable capacity is a latent state. It is revealed only by a discharge test, by an impedance or conductance test whose internal resistance rises as the battery ages, by an actual outage, or by a runtime estimate from a model fed with those test results. A latent state does not produce residuals until it is exercised, so its diagnosis is a matter of scheduled tests and of models of aging rather than of monitoring.

13.4.6 Generator start degradation

Mechanism. The time a standby generator takes to start and accept load, t_{start}, drifts upward through a weak starting battery, a fuel-system problem, starter wear or a control delay. In an illustrative series of monthly tests it reads 8, 12 and 18 s. The generator still starts every time, so every test passes.

Trend. The successive increases are 4 and 6 s. A quadratic through the three points, $t_{\text{start}} = 8 + 3n + n^2$ with n the number of months since the first test, gives 26 s at the next test; a straight line fitted to the three points by least squares, $t_{\text{start}} = 7.67 + 5n$, gives 23 s. Against an illustrative ride-through budget of 30 s, the time the UPS can carry the load before the generator must accept it (short with a flywheel UPS, and consumed quickly by failed start attempts with a battery UPS), the quadratic reaches the budget at $n = 3.4$ months and the line at $n = 4.5$ months.

Figure 13.4 plots the series. The horizontal axis is months since the first test; the vertical axis is the start time in seconds. The three dots are the tests; the dashed curve is the quadratic extrapolation and the dotted line the linear one; the red horizontal line is the ride-through budget. A binary record, "started" or "failed to start", shows three passes and no warning. The trend shows the margin closing within a few months. Diagnosis applies to degradation trends, not only to binary failures.

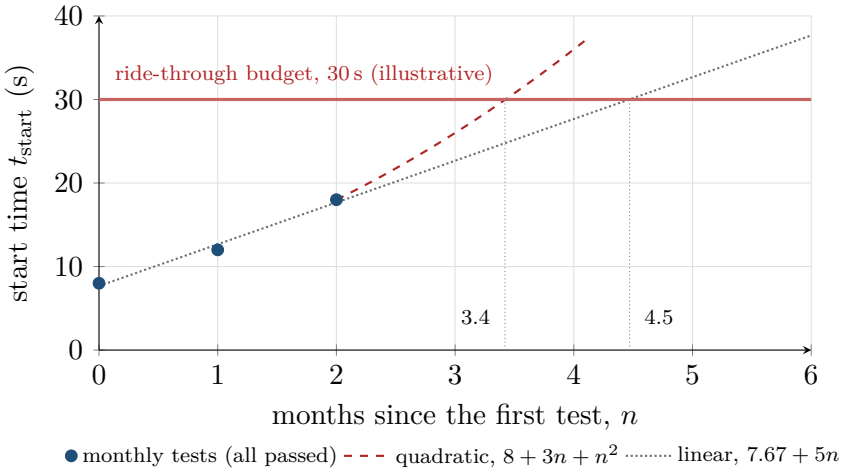


Figure 13.4: Generator start time in successive monthly tests, illustrative (8, 12, 18 s), with a quadratic extrapolation $t_{\text{start}} = 8 + 3n + n^2$ and the least-squares line $7.67 + 5n$, against an illustrative ride-through budget of 30 s. Every test passes, yet both extrapolations reach the budget within five months (at $n = 3.4$ and 4.5).

13.5 Residual-based diagnosis and ranking

The catalogue gives, for each failure mode, a predicted signature. This section turns the comparison between predicted and measured signatures into a procedure: residuals that measure disagreement, patterns of residuals that locate it, probabilities that rank the candidates, and the evidence in time and space that sharpens the ranking.

13.5.1 Residuals

A model of normal operation predicts each measurement from the estimated state $\hat{x}$ and the parameters θ,

$$\hat{y} = h(\hat{x}, \theta), \qquad r = y - \hat{y}, \tag{13.2}$$

and the residual r is what the model fails to explain. A residual near zero, within the noise of the measurement, says that the measurement is consistent with the model. A large residual says that something is inconsistent: the system, the sensor or the model has departed from what was assumed. Residuals are compared with their noise by normalizing, r/σ, where σ is the standard deviation of the residual under normal operation; a common working threshold is $|r|/\sigma > 3$.

One residual detects; it rarely identifies. A residual of $+3\,\mathrm{K}$ on a rack inlet temperature is consistent with all six hypotheses of Section 13.1.2. The value of residuals lies in having many of them, on different variables built from different physical relations, so that each hypothesis predicts a different pattern across them.

13.5.2 Residual patterns

Collect four residuals of one rack into a vector,

$$r = \left[\, r_T,\ r_P,\ r_{\dot{m}},\ r_{\Delta p} \,\right]^{\mathsf{T}}, \tag{13.3}$$

on temperature, rack power, airflow and filter pressure difference, each measured against a normal-operation prediction at the rack's usual operating point. Table 13.1 lists the pattern each of four hypotheses predicts. The rows are hypotheses and the columns residuals; each entry is the predicted direction. A high IT load raises temperature and power and leaves flow and pressure alone. A blocked filter raises temperature and filter pressure and lowers flow at unchanged power. A temperature sensor bias raises the temperature residual alone. A fan degradation raises temperature and lowers flow, like the blocked filter, but lowers the filter pressure difference, which falls with the flow through an unchanged filter. The table is the basis of model-based fault diagnosis: each row is a template, and the measured vector is matched against all of them.

Table 13.1: Residual patterns of four hypotheses for one rack: the predicted direction of each residual of equation (13.3) against the normal-operation prediction at the usual operating point.

Hypothesis	r_T	r_P	$r_{\dot{m}}$	$r_{\Delta p}$	separated from the others by
High IT load	high	high	normal	normal	r_P
Blocked filter	high	normal	low	high	$r_{\Delta p}$ high
Temperature sensor bias	high	normal	normal	normal	r_T alone
Fan degradation	high	normal	low	low	$r_{\Delta p}$ low

Two remarks fix how the table is read. First, the pattern depends on what the model takes as input. The residuals of Table 13.1 are measured against the usual operating point, so a load increase shows as a temperature residual. A model that takes the measured rack power as

an input predicts the higher temperature that the higher load causes, and under that model a load increase leaves r_T near zero; this is the construction of the context-aware alarms of Section 13.6.2. Second, the same idea extends to every failure mode of the catalogue. Table 13.2 collects the signatures of Sections 13.2 to 13.4: for each failure mode, its footprint, its signature and the evidence that separates it from its nearest rivals.

Table 13.2: Signatures of the failure modes of this chapter: footprint, signature, and the evidence that separates each from the failure modes it resembles.

Failure mode	Footprint	Signature	Distinguished by
High IT load (reference)	rack	P_{rack} up, then T_{out}, then possibly T_{in}	rack power rises first
Server fan degradation	one server	server exhaust and CPU temperature up, other fans faster, inlet normal	local; inlet and rack power normal
Blocked airflow	rack	ΔT_{rack} up, N_{fan} up, Δp_{filter} up, P_{rack} nearly unchanged	pressure across obstruction up; power not up
Hot-air recirculation	rack and neighbors	T_{in} and T_{out} up together, supply and power normal, hot aisle locally up	spatial gradient; ΔT_{rack} unchanged
CRAH fan degradation	row	row inlets up, cold-aisle pressure low, water normal	air side, not water side
Stuck chilled-water valve	one CRAH	supply air up, command high, coil water ΔT up	command and effect disagree; air side confirms
Pump degradation	loop, several CRAHs	several valves open, return water warmer, loop pressure low	all CRAHs on one loop
Chiller degradation	facility or chiller share	COP down at normal lift; $T_{\text{CHW,s}}$ up at high load	COP low for its lift
Cooling-tower degradation	facility	$T_{\text{CW,s}}$ up, lift up, COP down	approach high at measured T_{wb}
Temperature sensor bias	one sensor	one reading up, all related variables unchanged	physically inconsistent reading
Flow sensor bias	one meter	flow reading down, water ΔT and load unchanged	$\dot{Q}_{\text{flow}} \neq \dot{Q}_{\text{expected}}$
Topology error	rows involved	responses appear in the wrong row	effect moves rather than weakens
UPS degradation	electrical room	$P_{\text{in}} - P_{\text{out}}$ up, room warmer	loss up at unchanged output
Battery degradation	UPS	none in float operation	only tests, outages, runtime models
Generator start degradation	backup path	t_{start} rising across passing tests	trend against the ride-through budget

13.5.3 Overlap and measurement adequacy

Two hypotheses *overlap* on a set of measurements when they predict the same values for all of them. Fan degradation and a blocked filter both lower the airflow, so on the airflow residual alone they are indistinguishable. Separating them needs a measurement on which they differ: the fan speed N_{fan}, which rises under a blockage as the fans fight the resistance, or the filter pressure difference Δp_{filter}, which rises under a blockage and falls under a weak fan. Whether two hypotheses can be separated is a property of the measurement set, not of the reasoning applied to it; this is *measurement adequacy*.

Figure 13.5 shows the overlap geometrically. The horizontal axis is the normalized airflow

residual $r_{\dot{m}}/\sigma$ and the vertical axis the normalized filter pressure residual $r_{\Delta p}/\sigma$. Each ellipse is the region in which the residuals of one hypothesis fall, about two standard deviations around its predicted center: the blocked filter at upper left, the fan degradation at lower left, and at the origin the hypotheses that leave flow and pressure alone, normal operation, high IT load and temperature sensor bias. The colored bars along the bottom edge are the projections of the ellipses on the airflow axis alone; the bars along the left edge are their projections on the pressure axis. The star is an illustrative measurement. On the airflow axis the two left bars coincide, and the measurement cannot choose between a blockage and a weak fan. On the pressure axis the bars are separate, and the measurement falls in the blocked-filter bar. The three hypotheses at the origin overlap in this plane entirely, and are separated by the other two residuals, r_T and r_P.

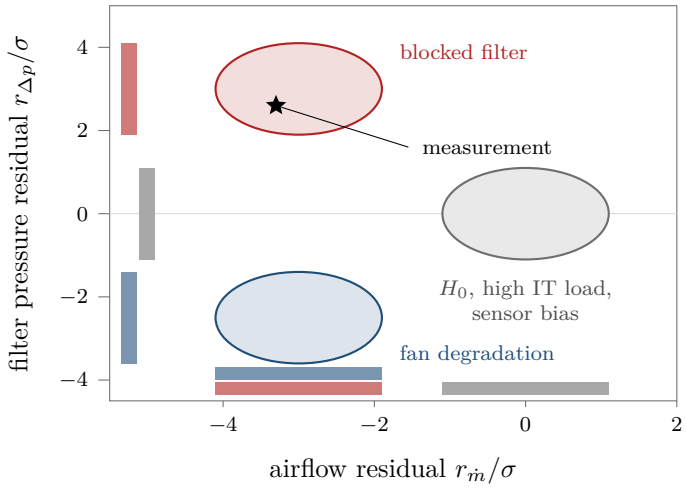


Figure 13.5: Overlap of hypotheses in the plane of two normalized residuals, illustrative. Ellipses mark the regions (about two standard deviations) predicted by each hypothesis; bars along the bottom and left edges are their projections on each axis. On airflow alone a blocked filter and a fan degradation coincide; the filter pressure separates them.

13.5.4 Probabilistic ranking

Real residuals rarely fall cleanly inside one template, and a hard classification, "the cause is a blocked filter", discards the information that a second explanation is nearly as good. The more honest and more useful output is a probability for each hypothesis given the measurements, $\Pr(H_i \mid y)$, such as

$$\text{blocked airflow } 0.45, \quad \text{fan degradation } 0.30, \quad \text{sensor bias } 0.15, \quad \text{high IT load } 0.10.$$

Bayes' rule produces it from two ingredients,

$$\Pr(H_i \mid y) = \frac{\Pr(y \mid H_i)\,\Pr(H_i)}{\sum_j \Pr(y \mid H_j)\,\Pr(H_j)} \;\propto\; \Pr(y \mid H_i)\,\Pr(H_i), \tag{13.4}$$

the prior $\Pr(H_i)$ and the likelihood $\Pr(y \mid H_i)$. The physics model defines the likelihood: it says what each hypothesis predicts and therefore how probable the observed values are under it. Under a high IT load the rack power should rise; if it does not, the measured power is improbable under H_1, $\Pr(y \mid H_1)$ is small, and the posterior of H_1 falls whatever its prior.

Figure 13.6 works the ranking above through equation (13.4). The horizontal axis lists the four hypotheses retained for one rack; for each there are three bars, the prior, the likelihood of the measurements, and the posterior. The illustrative priors are 0.30, 0.20, 0.30 and 0.20;

the illustrative likelihoods are 0.6, 0.6, 0.2 and 0.2. The products are 0.18, 0.12, 0.06 and 0.04, which sum to 0.40; dividing by the sum gives the posteriors 0.45, 0.30, 0.15 and 0.10. Sensor bias starts with the same prior as blocked airflow and ends with a third of its posterior, because the measurements are three times less probable under it. The probabilistic diagnosis of Chapter 16 develops how such likelihoods are computed from the physics model and the noise of each sensor.

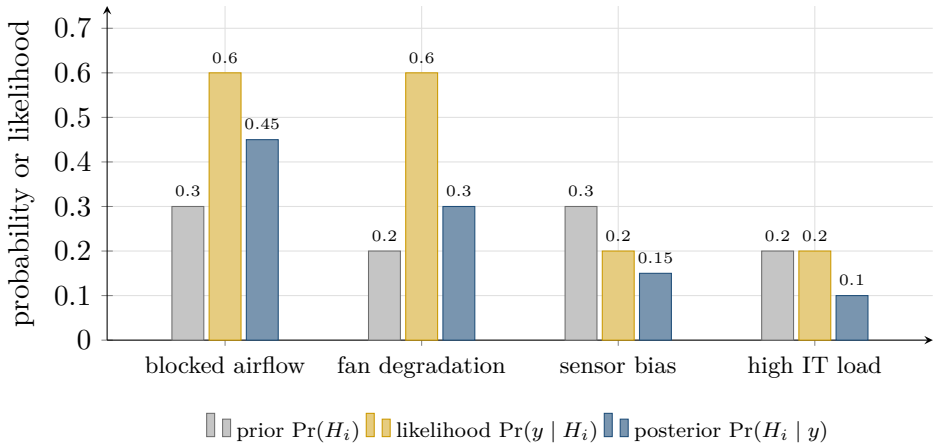


Figure 13.6: Bayes' rule, equation (13.4), for four hypotheses of one rack, illustrative. Posterior = prior × likelihood divided by $\sum_j \Pr(y \mid H_j) \Pr(H_j) = 0.40$. The ranking moves from the priors toward the hypotheses under which the measurements are most probable.

Priors. The prior carries the failure history of the equipment. A temperature sensor bias and a complete failure of a chiller compressor may explain a given set of sparse readings equally well, so that their likelihoods are equal. Their priors are not: sensor biases occur far more often than compressor failures. With illustrative monthly probabilities of 2 % for a bias in a given sensor and 0.1 % for a compressor failure, the prior odds are 20 to 1, and with equal likelihoods the posterior odds are the same 20 to 1 in favor of the sensor. Diagnosis combines three things: the physics, which sets the likelihoods; the evidence, which selects among them; and the prior reliability of each component, which Chapter 12 supplies.

13.5.5 Temporal and spatial evidence

Order in time is evidence. A cause precedes its effects, and the physical chain from a fault to a temperature takes time. Suppose a rack inlet temperature starts to rise at 10:00. If the rack power rose at 9:57, the timing supports a load increase: the power changed first and the temperature followed. If instead the chilled-water supply temperature rose at 9:55, the timing supports a plant problem: the warmer water reached the coil, the supply air warmed, and the rack followed.

Figure 13.7 draws the two cases. Each panel shows three signals against clock time from 9:50 to 10:10, each normalized so that 0 is its value before the event and 1 its final change. Panel (a) is the load case: rack power steps at 9:57 and the inlet temperature follows from 10:00, while the chilled-water supply does not move. Panel (b) is the plant case: the chilled-water supply rises from 9:55, the inlet temperature follows from 10:00, and the rack power does not move. At 10:05 the inlet temperature is the same in both panels; only the other two signals and their order separate the causes. Time-aligned records are a precondition: a clock offset of a few minutes between the power meters and the plant controllers erases the evidence (Chapter 9).

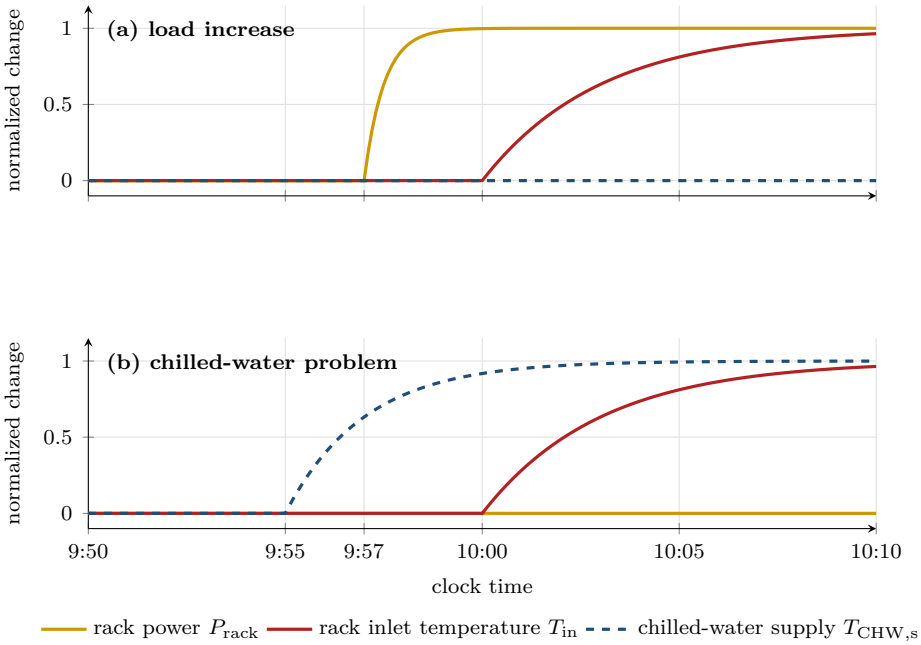


Figure 13.7: Temporal evidence, illustrative. Each signal is normalized to its total change. (a) Load increase: rack power rises at 9:57 ($1 - e^{-(t-t_0)/0.5\,\text{min}}$), inlet temperature at 10:00 ($1 - e^{-(t-t_0)/3\,\text{min}}$). (b) Chilled-water problem: supply water rises at 9:55 ($1 - e^{-(t-t_0)/2\,\text{min}}$), inlet temperature at 10:00. The inlet temperature is identical in both panels; the order of the other signals identifies the cause.

Spatial evidence. Location is evidence in the same way, and Figure 13.3 already gave the rule. A problem confined to one rack points to a local fan, a local obstruction, local recirculation or a sensor. A problem across a whole row points to a CRAH, a pump branch or a control zone. A problem across the whole facility points to the chiller plant, the utility supply, a global control setting or the weather. A diagnosis that compares each rack with its neighbors, each row with the others and each loop with its siblings uses this evidence; a diagnosis that looks at each rack alone does not.

13.5.6 Fault propagation and competing causal explanations

A fault does not change the measurements directly. It changes a physical state, which changes other states, which change what the sensors read:

$$\text{fault} \rightarrow \text{physical state changes} \rightarrow \text{sensor changes.}$$

A blocked filter raises the flow resistance; the rack airflow $\dot{m}_{\text{air}}$ falls; the rack temperature rise ΔT_{rack} grows; the processor temperature T_{CPU} rises. This causal chain is a *fault propagation graph*: nodes are states, arrows are physical relations, and the order along the arrows is the order in which the observations change. The variables nearest the fault change first and most; those farther along change later and are mixed with the effects of other causes.

The same final symptom can be reached by several chains. A rising processor temperature can come from a higher processor power P_{CPU}, which raises the heat $\dot{Q}_{\text{CPU}}$ the chip must shed (H_1); from a lower airflow, which lowers the heat the air removes (H_2); or from a biased processor temperature sensor, with no physical change at all (H_3). Figure 13.8 draws the three chains. Each row is one hypothesis, read from left to right; the gray boxes are physical states, and the dashed boxes are the measured variables that each chain changes on its way to the symptom. The chains converge on the true processor temperature (first two) or on the reading

alone (third). The final symptom does not distinguish them; the intermediate consequences do. Rack power rises only under the first chain, fan speed, filter pressure and exhaust temperature move only under the second, and nothing else moves under the third.

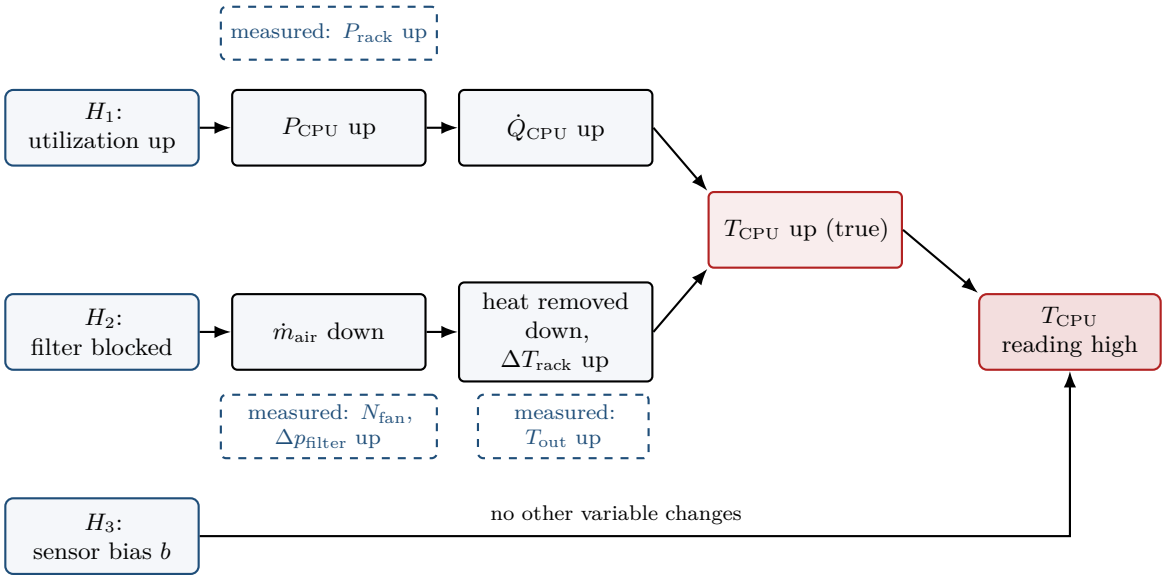


Figure 13.8: Three causal chains to one symptom, a high processor temperature reading. Gray boxes are physical states, dashed boxes the measured variables each chain changes on the way. Only the intermediate consequences, rack power for H_1 and fan speed, filter pressure and exhaust temperature for H_2, distinguish the chains; H_3 changes the reading alone.

13.5.7 Detection, isolation, identification and severity

Diagnosis has three levels of ambition. *Detection* decides whether there is a fault at all. *Isolation* decides which subsystem it is in: the air side, the water side, the electrical system or the sensors. *Identification* decides exactly what the fault is and how large, for example a server fan degraded by 30 %. Each level needs more than the one before, in measurements and in model, as Table 13.3 lists: its rows are the three levels, and its columns the question, the output and what the level requires.

Table 13.3: The three levels of diagnosis and what each requires.

Level	Question	Output	Requires
Detection	is there a fault?	H_0 or not H_0	one residual with a known noise level
Isolation	in which subsystem?	air, water, electrical or sensor	residuals in several domains: temperature, power, flow, pressure
Identification	what exactly, and how large?	a fault and its size, e.g. fan $-30\,\%$	measurements sensitive to the fault parameter, and a model that contains it

Severity. Identification includes the size of the fault. An airflow loss of 5 % and one of 50 % are the same failure mode with very different consequences: at constant heat the rack temperature rise scales as $1/\dot{m}_{\text{air}}$, so a normal 12 K rise becomes $12/0.95 = 12.6$ K in the first case and $12/0.5 = 24$ K in the second. The size is a parameter, θ_{block}, the fractional loss of

airflow, and it is estimated with its uncertainty as a distribution $p(\theta_{\text{block}} \mid y)$. For an illustrative rack at $P_{\text{rack}} = 18\,\text{kW} \pm 2\,\%$ with a normal airflow of $1.5\,\text{kg/s}$ and a measured rise of $15 \pm 0.5\,\text{K}$, the calculated airflow is $18/15 = 1.2\,\text{kg/s}$, so $\theta_{\text{block}} = 1 - 1.2/1.5 = 0.20$. Its relative uncertainty combines the two relative errors, $\sqrt{0.02^2 + (0.5/15)^2} = 0.039$, so the airflow is $1.2 \pm 0.047\,\text{kg/s}$ and $\theta_{\text{block}} = 0.20 \pm 0.03$. The distribution, not only the point value, is what feeds the risk prediction of the next section.

Principle 13.4. *One residual detects a fault; a pattern of residuals across domains isolates it; measurements sensitive to the fault's parameter identify it and its size. Each level needs more instrumentation than the last.*

13.6 From diagnosis to action

A ranked diagnosis is an intermediate product. This section connects it to what an operator does: which hypothesis to check first, how alarms should be built, what to conclude when faults combine or measurements are few, which measurement to take next, and how the whole sequence is organized.

13.6.1 Risk-weighted priority

The most probable hypothesis is not necessarily the one to act on first. Action is ordered by the expected loss of each hypothesis, the product of its probability and its consequence,

$$\mathcal{R}_i = \Pr(H_i \mid y)\, L_i, \tag{13.5}$$

where L_i is the loss if H_i is true and goes unaddressed. A hypothesis of moderate probability and severe consequence can outrank a likely but harmless one.

Table 13.4 works an illustrative case. Its rows are three hypotheses for one warm row; its columns are the posterior, the loss if the hypothesis is true and ignored, and their product. A sensor bias has probability 0.60 but negligible immediate consequence: a misleading reading that costs a technician visit, about \$2,000 in the illustrative units. A cooling failure has probability 0.25 but a severe thermal consequence, throttling or shutdown of the row if it progresses, about \$400,000. The expected losses are \$1,200 and \$100,000. The cooling failure is checked first, even though the sensor bias is more than twice as probable. This is risk-based diagnosis, and Chapter 17 develops risk as a probability of insufficient capacity.

Table 13.4: Risk-weighted priority for one warm row, illustrative: posterior, loss if true and unaddressed, and expected loss, equation (13.5).

Hypothesis	$\Pr(H_i \mid y)$	loss L_i (\$)	$\mathcal{R}_i$ (\$)	priority
Temperature sensor bias	0.60	2,000	1,200	3
Cooling failure (CRAH or plant)	0.25	400,000	100,000	1
High IT load	0.15	10,000	1,500	2

Principle 13.5. *Act on probability times consequence, not on probability alone: a moderately probable cause with a severe consequence is verified before a likely cause with a negligible one.*

13.6.2 Residual and context-aware alarms

A fixed threshold, $T > 28°\mathrm{C}$, fires whenever the temperature crosses it, including when the rise is exactly what the operating conditions predict. If the rack power has just risen, a warmer rack is expected, and an alarm on it is a false alarm. The remedy is to alarm on the residual instead of on the value: predict the temperature from the operating context,

$$T_{\text{expected}} = f(P_{\text{IT}}, \dot{m}, T_{\text{supply}}), \qquad r_T = T_{\text{obs}} - T_{\text{expected}}, \tag{13.6}$$

and alarm when $|r_T|$ is large compared with its noise.

The same value then means different things in different contexts. Take an illustrative inlet model fitted to normal operation, $T_{\text{expected}} = T_{\text{supply}} + \kappa\, P_{\text{rack}}$ with $T_{\text{supply}} = 21°\mathrm{C}$ and $\kappa = 0.1\,\mathrm{K/kW}$, the inlet rising with rack power through recirculation, and a residual noise of $\sigma = 0.8\,\mathrm{K}$. When the rack power rises from 50 to 75 kW, the expected inlet goes from 26 to 28.5°C; a reading of 28.6°C trips the fixed 28°C threshold, but its residual is 0.1 K and the residual alarm stays quiet. A reading of 29°C is normal at $P_{\text{rack}} = 80\,\mathrm{kW}$, where the expectation is 29°C, and suspicious at 20 kW, where the expectation is 23°C and the residual of 6 K is 7.5σ. A context-aware alarm reports the state relative to its operating context, not relative to a fixed limit; the fixed limit remains as a safety bound on the physical temperature.

13.6.3 Multiple simultaneous faults

Faults do not always arrive one at a time. A rack can carry a high IT load and a partially blocked air path together. The pattern is then roughly the sum of the two single-fault patterns of Table 13.1: r_T high, r_P high, $r_{\dot{m}}$ low and $r_{\Delta p}$ high, which matches neither single-fault template. A diagnosis restricted to single faults will force it into the nearest one and be wrong about the other. The joint hypothesis needs its own posterior, $\Pr(H_1, H_2 \mid y)$.

Allowing combinations is combinatorially harder. The fifteen failure modes of Table 13.2 give $15 \times 14/2 = 105$ pairs, 455 triples and $2^{15} = 32{,}768$ subsets in all. The prior restrains the search: if each fault has an illustrative independent monthly probability of 0.01, a given pair has 10^{-4}, so multiple-fault hypotheses need much stronger evidence than single ones and are examined when no single fault explains the data. They are rarer, but they are realistic, most often when one fault has persisted unnoticed and a second arrives.

13.6.4 Sparse measurements and the next measurement

Suppose only the rack inlet temperature and the rack power are measured. A temperature rise at constant power says that something thermal changed, and it rules out a load increase. It does not separate a server fan degradation, recirculation and a warming of the CRAH supply: all three raise the inlet, or the temperature read near it, at constant power, and the two available measurements predict the same values under each. The correct conclusion is then "insufficient information to isolate the fault". It is a valid diagnostic result, not a failure of the diagnosis, and it is more useful than a confident wrong answer, because it says what is missing.

What is missing is chosen as the *next-best measurement*: the measurement whose predicted values differ most between the leading hypotheses. With fan degradation and recirculation in the lead, three candidates are the server fan speed, the pressure difference across the rack, and the hot-aisle temperature behind it. Table 13.5 lists, for each candidate, the change it is predicted to show under each hypothesis, its noise σ, and the separation $d = |\mu_1 - \mu_2|/\sigma$ between the two predicted means (Chapter 9). The fan speed separates them best, $d = 3.8$; the hot-aisle temperature separates them worst, $d = 1.4$, because recirculation raises the inlet and the exhaust together and so warms the hot aisle almost as much as a weak fan does. Chapter 19 develops the choice as an expected information gain.

Table 13.5: Candidate next measurements for separating server fan degradation from recirculation, illustrative: predicted change under each hypothesis, noise, and separation $d = |\mu_1 - \mu_2|/\sigma$.

Candidate measurement	fan degradation	recirculation	σ	d
Speed of the server's other fans (%)	+25	+6	5	3.8
Pressure difference across the rack (Pa)	−20	0	8	2.5
Hot-aisle temperature behind the rack (K)	+5.4	+4.0	1.0	1.4

Principle 13.6. *"Insufficient information to isolate the fault" is a valid diagnostic conclusion. It names the hypotheses that remain and leads to the measurement that would separate them.*

13.6.5 Recommendation and the diagnostic workflow

A diagnosis ends with an action, stated with the evidence that supports it. Two illustrative recommendations show the form. "Most likely cause: CRAH airflow restriction. Check the filter differential pressure and the fan status." Or, when the evidence does not separate the leading hypotheses: "Current evidence cannot distinguish a valve fault from a flow-sensor bias. Verify the branch flow before replacing the valve." The second is the more valuable of the two in practice, because it prevents the replacement of a working valve.

Figure 13.9 draws the full sequence. The boxes are seven steps, numbered in order, running left to right along the upper row and right to left along the lower row; the arrow from the last step back to the first is the loop that a new measurement or an action closes. The steps are: (1) detect an inconsistency, a residual larger than its noise; (2) generate the plausible hypotheses from the catalogue and the footprint; (3) predict the signature of each with the physics model; (4) compare the predictions with the evidence; (5) rank the hypotheses by posterior probability; (6) assess the consequence of each; (7) recommend an action or the next measurement. The loop matters: the measurement recommended at step 7 is new evidence at step 4, and the diagnosis tightens with each pass.

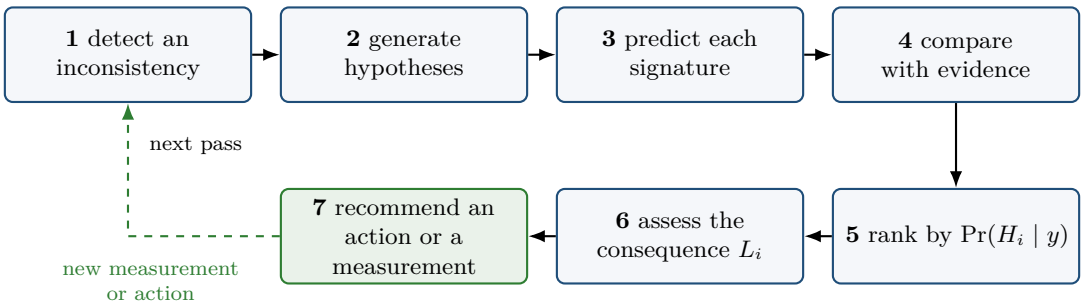


Figure 13.9: The seven-step diagnostic workflow. Steps 1 to 4 find and test explanations, steps 5 and 6 rank and weigh them, step 7 turns them into an action or a measurement, whose result starts the next pass.

13.6.6 A joint probabilistic formulation

The steps of the workflow are pieces of one probabilistic object, the joint posterior over the fault hypothesis H and the hidden physical state x given the measurements y,

$$p(H, x \mid y) \;\propto\; p(y \mid x, H)\, p(x \mid H)\, \Pr(H). \tag{13.7}$$

Summing over the states gives the ranking of hypotheses, $\Pr(H_i \mid y) = \int p(H_i, x \mid y)\,\mathrm{d}x$; summing over the hypotheses gives the estimate of the state, $p(x \mid y) = \sum_i p(H_i, x \mid y)$, which is the subject of state estimation; and pushing that state forward through the physics gives the probability of a thermal violation,

$$\Pr(\text{violation} \mid y) = \sum_i \Pr(H_i \mid y)\ \Pr(T_{\text{in}} > T_{\max} \mid H_i, y). \tag{13.8}$$

With the posteriors of Figure 13.6 and illustrative conditional violation probabilities of 0.30 under blocked airflow, 0.20 under fan degradation, 0.01 under sensor bias and 0.10 under a high IT load, equation (13.8) gives $0.45 \times 0.30 + 0.30 \times 0.20 + 0.15 \times 0.01 + 0.10 \times 0.10 = 0.135 + 0.060 + 0.0015 + 0.010 \approx 0.21$. The same object yields the diagnosis, the state and the risk; diagnosis and risk are unified.

Beyond anomaly detection. Three levels of statement are possible about abnormal telemetry. Anomaly detection says that a behavior is unusual. Diagnosis says why. Physical diagnosis says which explanations are consistent with the conservation laws and the topology of the system, and with what probability. The first needs only the history of the data; the second needs a list of causes; the third needs the physical model, which is what makes its explanations checkable against every other measurement in the facility.

13.7 What to retain

Definition 13.1 (Physical diagnosis)**.** A symptom is not a fault. Load, airflow, water flow, equipment degradation, topology and sensor error are competing explanations of the same symptom; diagnosis asks not which value crossed a threshold but which physical explanation best accounts for all the available evidence, ranked by $\Pr(H_i \mid y) \propto \Pr(y \mid H_i)\Pr(H_i)$ and acted on by probability times consequence.

Table 13.6 lists the chapter's objects and the subsection that defines each.

Table 13.6: The objects of failure diagnosis and where each is defined.

Object	Expression	Role	Defined in
Hypotheses	$H_0, H_1, \ldots, H_6$	competing explanations	13.1.2
Signature	pattern across variables	what each cause predicts	13.1.3
Blocked airflow	$\Delta p = K\dot{V}^2$, Figure 13.2	operating point shifts	13.2.2
Commanded and actual	command high, effect absent	actuator fault evidence	13.3.1
Footprint	Figure 13.3	locates a fault on the chain	13.3.5
Tower approach	$T_{\text{CW,s}} - T_{\text{wb}}$	tower versus chiller	13.3.4
Consistency check	$P_{\text{rack}} = \dot{m}_{\text{air}} c_{p,\text{air}}(T_{\text{out}} - T_{\text{in}})$	exposes a biased sensor	13.4.1
Degradation trend	$t_{\text{start}}(n)$	margin before failure	13.4.6
Residual	$r = y - h(\hat{x}, \theta)$	detects inconsistency	13.5.1
Residual pattern	$[r_T, r_P, r_{\dot{m}}, r_{\Delta p}]^{\mathsf{T}}$	isolates the cause	13.5.2
Overlap	Figure 13.5	measurement adequacy	13.5.3
Posterior ranking	equation (13.4)	likelihood times prior	13.5.4
Propagation graph	fault → state → sensor	order of observations	13.5.6
Severity	$p(\theta_{\text{block}} \mid y)$	size of the fault	13.5.7
Expected loss	$\mathcal{R}_i = \Pr(H_i \mid y) L_i$	priority of action	13.6.1
Residual alarm	$r_T = T_{\text{obs}} - T_{\text{expected}}$	context-aware alarm	13.6.2
Next measurement	$d = \lvert\mu_1 - \mu_2\rvert/\sigma$	separates leading hypotheses	13.6.4
Joint posterior	$p(H, x \mid y)$	diagnosis, state and risk	13.6.6

13.8 Exercises

Exercise 13.1. The server fans of a rack act as one equivalent fan with the curve $\Delta p = 500 - 125\,\dot{V}^2$ (Pa, m^3/s) at nominal speed, and the clean rack path has $K = 200\,\mathrm{Pa}/(\mathrm{m}^3/\mathrm{s})^2$. (a) Find the operating flow and pressure. (b) A blockage doubles K. Find the new flow at nominal speed and the percentage loss. (c) With an air density of $1.2\,\mathrm{kg}/\mathrm{m}^3$ and a rack heat of 15 kW, compute the rack temperature rise before and after the blockage. (d) Using the scaled fan curve $\Delta p = 500\,s^2 - 125\,\dot{V}^2$, find the speed ratio s that restores the original flow and the factor by which fan power rises. Which measured variables distinguish this case from a load increase that produces the same temperature rise?

Exercise 13.2. A chiller removes 2 MW and draws 470 kW. The fraction-of-Carnot model has $\eta_C = 0.5$ and $T_{\mathrm{evap}} = 278\,\mathrm{K}$, and the normal lift is 27.8 K at a tower approach of 4 K. The wet-bulb temperature is 20°C and the condenser-water supply reads 27°C. (a) Compute the tower approach and the excess over design. (b) Assuming the condensing temperature follows the condenser water, compute the lift and the COP the model predicts, and the compressor power it implies for 2 MW. (c) Compute the measured COP. Is the tower degradation alone enough to explain the measured power, or is the chiller also degraded? By how many kilowatts?

Exercise 13.3. Four hypotheses for a warm rack have priors 0.25 (blocked airflow), 0.25 (fan degradation), 0.40 (sensor bias) and 0.10 (CRAH degradation), and likelihoods of the measurements 0.5, 0.3, 0.1 and 0.6. (a) Compute the posteriors with equation (13.4). (b) The losses if each is true and unaddressed are \$50,000, \$20,000, \$1,000 and \$300,000. Compute the expected losses and the order of action. Does it differ from the order of the posteriors? (c) With conditional violation probabilities 0.3, 0.2, 0 and 0.5, compute $\Pr(\text{violation} \mid y)$ with equation (13.8).

Exercise 13.4. (a) A chilled-water flow meter reads 6.5 kg/s while the water temperature rise across the coil is 5.5 K and the air-side balance gives 200 kW. Compute $\dot{Q}_{\mathrm{flow}}$, the residual against $\dot{Q}_{\mathrm{expected}}$, and the water temperature rise that a true flow of 6.5 kg/s would produce at 200 kW. Which explanation do the temperatures support? (b) A rack draws 16 kW; its exhaust reads 38°C and its inlet sensor 30°C. The rack's normal airflow, with fan speeds unchanged, is 1.6 kg/s ($c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ}/(\mathrm{kg\,K})$). Compute the airflow implied by the readings, the inlet temperature implied by the normal airflow, and the inferred sensor bias. Which other measurements would confirm the bias?

13.9 Connection to the TERRA framework

The failure modes of this chapter are departures from the conservation networks of the earlier chapters, and diagnosis is inference on those networks with the departure itself as an unknown.

- **Each failure mode is a parameter or state shift on a graph element.** A server or CRAH fan degradation lowers the curve of a fan component on an air edge; a blocked filter raises the resistance K of an air edge; a stuck valve separates the actual position of a valve component from its command; a fouled chiller or tower lowers the COP or raises the approach of a plant component; a degraded UPS lowers the efficiency of an electrical component whose loss enters the thermal network as a heat source. The catalogue of Table 13.2 is a list of such shifts.
- **Balances that should close are the residuals.** The rack energy balance, the coil balance of air against water, the chiller balance $\dot{Q}_{\mathrm{cond}} = \dot{Q}_{\mathrm{evap}} + W_{\mathrm{comp}}$ and the UPS input against output are conservation laws at nodes coupling the air, water, thermal and electrical networks; each residual of equation (13.2) is one of them failing to close, and the pattern of Table 13.1 is which ones fail.

- **Sensor bias and topology are unknowns of the same inference.** A bias b is a state attached to one measurement and is inferred from the balances it violates, as in Section 13.4.1; the assignment of CRAH 3 to row A or row B is an uncertain edge of the graph G, tested by the responses of Section 13.4.3.
- **The ranked diagnosis is a posterior over hypotheses.** Each hypothesis is a version of the graph model with one shift switched on; forward simulation gives its predicted signature, the likelihood compares it with the measurements, and the result is the ranking $\Pr(H_i \mid y)$ of Figure 13.6. The joint posterior $p(H, x \mid y)$ of equation (13.7) carries the inferred states, rack airflow, recirculation fraction, coil conductance, alongside.
- **Risk and the next measurement come from the same posterior.** Pushing $p(H, x \mid y)$ forward gives the probability of a thermal violation, equation (13.8), and the expected losses of Table 13.4; the candidate measurements of Table 13.5 are ranked by how much they would separate the leading hypotheses.

In summary, a fault is a physical departure at one element of the coupled power, air, water and thermal networks, it reaches the sensors through a causal chain whose order and footprint are themselves evidence, and diagnosis ranks the candidate departures by how well each explains all the measurements, how often each occurs and how much each would cost. The next chapter, Chapter 14, turns to the estimate of the hidden state x itself: how the measurements and the model are combined into $p(x \mid y)$.

Chapter 14

State Estimation

14.1 Measurements and the physical state

The earlier chapters described what a data center measures and how its sensors err (Chapter 9), and the chapter before this one used residuals to detect that something is wrong (Chapter 13). This chapter treats the step that both rest on: turning a set of readings into a statement about the physical state that produced them. The true state of the facility, every airflow, temperature, heat flow and valve position, is never observed directly; only measurements are. The question is

given incomplete and noisy measurements, what is the most plausible internal physical state of the data center, and how certain is that statement?

The answer is a method, state estimation, that combines three ingredients: a model of how the state evolves, a model of what each sensor reports, and the readings themselves. The chapter develops it in two forms, dynamic (a filter that runs as data arrive) and static (a weighted fit to one snapshot), applies it to the hidden quantities a data center cares about, states when the measurements are sufficient to determine them at all, and ends with what an estimate is used for and how it fails.

Jargon: state, parameter, state estimation, process noise, measurement noise, innovation, Kalman filter, Kalman gain, weighted least squares, observability, posterior, MAP estimate, filtering, smoothing

The *state* x of a system is the set of quantities that change on the time scale of operation and that, with the inputs, determine its future: temperatures, mass flows, pressures, valve positions. A *parameter* θ is a quantity that the model treats as constant or slowly varying, such as a thermal resistance or the conductance of a coil. *State estimation* is the inference of the state from measurements through a model. *Process noise* w_k is the random term that represents what the model of the state's evolution leaves out; *measurement noise* v_k is the random error of a reading. The *innovation* is the residual between a new reading and the value the model predicted for it before the reading arrived. The *Kalman filter* is the recursive estimator that is optimal for a linear model with normally distributed noise; its *Kalman gain* K_k sets how much of each innovation is added to the prediction. *Weighted least squares* (WLS) chooses the state that minimizes the sum of squared residuals, each weighted by the inverse of its variance. *Observability* is the property that the measurements determine the state uniquely. The *posterior* $p(x \mid y)$ is the probability density of the state given the measurements; the *maximum a posteriori* (MAP) estimate is the state at which it peaks. *Filtering* estimates the current state from the measurements received so far; *smoothing* estimates a past state from measurements before and after it.

Key idea

A data center does not see its state x; it sees measurements y. State estimation combines a physical model with the measurements into a belief about the state, $p(x \mid y)$, whose center is the estimate and whose width is its uncertainty. Diagnosis, risk and control all act on that belief, so its correctness depends on the model and on whether the measurements determine the state at all.

14.1.1 Measured and hidden quantities

A rack and its cooling zone illustrate the distinction. The rack inlet temperature T_{in}, the rack exhaust temperature T_{out}, the rack power P_{rack}, the pressure difference Δp between the underfloor plenum and the room, and the chilled-water supply temperature $T_{\text{CHW,s}}$ are measured: a sensor reports each, with an error. The mass flow of air through the rack $\dot{m}_{\text{rack}}$, the recirculation fraction α at its inlet, the heat removed by the cooling unit $\dot{Q}_{\text{cool}}$, the overall conductance UA of the cooling coil, and the true position of the chilled-water valve are not. They are hidden states and parameters, and they are often the quantities an operator needs: whether the rack receives enough air, whether hot exhaust is reaching its inlet, whether the coil is fouling, whether the valve does what it is told.

The measurements and the state are therefore different objects, $y \neq x$. Some measured quantities are components of the state, such as a temperature at the sensor's location; others are functions of several states, such as a pressure difference or an electrical power that is the sum of many servers. Most of the state is not measured at all. The work of state estimation is to fill in the hidden components from the measured ones, using the physical relations that connect them, and to say how well the filling-in is determined.

14.1.2 The estimation problem

The connection between state and measurement is the measurement equation of Chapter 9,

$$y = h(x, \theta) + \epsilon, \tag{14.1}$$

where h is the function of the state and parameters that the sensors respond to and ϵ is the measurement error. State estimation inverts this relation: given y, it infers x. The inversion is rarely exact. The function h usually has fewer outputs than x has components, the error ϵ blurs every reading, and h itself is a model that may be wrong.

The probabilistic statement of the problem replaces "the" answer by a distribution. Bayes' rule gives

$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)} \;\propto\; p(y \mid x)\, p(x), \tag{14.2}$$

where the likelihood $p(y \mid x)$ follows from equation (14.1) and the distribution of ϵ, the prior $p(x)$ states what was known about the state before the readings (physical bounds, the previous estimate, the model's prediction), and $p(y)$ is a normalizing constant. The posterior $p(x \mid y)$ is the complete answer: its center is the most plausible state, its width is the uncertainty, and its shape shows whether more than one state is plausible. Every method in the chapter, from a quotient of three readings to a Kalman filter, computes this posterior or a summary of it.

14.1.3 Rack airflow as an inferred state

The simplest useful estimate is the airflow through one rack. The energy balance of Chapter 5, $\dot{Q} = \dot{m} c_p \Delta T$, applied to the air passing through the rack, together with the assumption that the rack's electrical power leaves as heat in that air, $\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$, gives

$$\hat{\dot{m}}_{\text{rack}} = \frac{P_{\text{rack}}}{c_{p,\text{air}}\,(T_{\text{out}} - T_{\text{in}})}. \tag{14.3}$$

For an illustrative rack with $P_{\text{rack}} = 30\,\text{kW}$, $T_{\text{in}} = 22°\text{C}$ and $T_{\text{out}} = 37°\text{C}$, the temperature rise is $\Delta T = 15\,\text{K}$ and, with $c_{p,\text{air}} \approx 1.0\,\text{kJ/(kg K)}$,

$$\hat{\dot{m}}_{\text{rack}} = \frac{30\,\text{kW}}{1.0\,\text{kJ/(kg K)} \times 15\,\text{K}} = 2.0\,\text{kg/s}.$$

The three inputs are measured; the formula is a calculation; the output is an inference about a quantity no sensor reports. The distinction matters because the inference carries the assumptions of the formula: that no heat leaves by another path (no liquid cooling, negligible conduction to the rack structure), that the rack is in steady state so that no heat is being stored, and that the two temperature points represent the mixed inlet and the mixed exhaust. If any assumption fails, the number 2.0 kg/s is still computed, and it is wrong.

14.1.4 Uncertainty of the airflow estimate

A point estimate without its uncertainty cannot be used to make a decision, because it does not say whether 2.0 kg/s is distinguishable from 1.7 kg/s. Suppose the rack power is metered to $P_{\text{rack}} = 30 \pm 1\,\text{kW}$ and each temperature sensor is accurate to $\pm 0.5\,\text{K}$ (one standard deviation, independent errors). The temperature difference then has a standard deviation of

$$\sigma_{\Delta T} = \sqrt{0.5^2 + 0.5^2} = 0.71\,\text{K},$$

4.7 % of 15 K. For a quotient the relative standard deviations add in quadrature, so

$$\frac{\sigma_{\hat{\dot{m}}}}{\hat{\dot{m}}} = \sqrt{\left(\frac{1}{30}\right)^2 + \left(\frac{0.71}{15}\right)^2} = \sqrt{0.033^2 + 0.047^2} = 0.058,$$

and $\hat{\dot{m}}_{\text{rack}} = 2.0 \pm 0.12\,\text{kg/s}$. The temperature difference contributes more than the power: a small ΔT is measured with a large relative error, and the airflow estimate inherits it.

The two sensor errors are not the only ones. A single exhaust sensor reads one point of an exhaust stream that is often stratified, so its reading differs from the mixed exhaust temperature by an amount that is not a sensor error but a representativeness error. Adding an illustrative 0.7 K for it raises $\sigma_{\Delta T}$ to $\sqrt{0.5^2 + 0.5^2 + 0.7^2} = 0.99\,\text{K}$ and the relative uncertainty to $\sqrt{0.033^2 + 0.066^2} = 0.074$, which gives $\hat{\dot{m}}_{\text{rack}} = 2.0 \pm 0.15\,\text{kg/s}$. That is the form in which the estimate should be reported: a center and a width, or, when the errors are not normal or the relation is strongly nonlinear, the full posterior $p(\dot{m}_{\text{rack}} \mid y)$, computed by sampling the inputs and evaluating equation (14.3) for each sample.

14.1.5 Redundant evidence

The energy balance is not the only relation that involves the rack airflow. Across a fixed flow restriction, such as the rack's perforated door or its filter, the pressure difference and the volumetric flow are related by $\Delta p = K\dot{V}^2$ (Chapter 6), so

$$\dot{V} = \sqrt{\Delta p / K}, \qquad \dot{m} = \rho\,\dot{V}, \tag{14.4}$$

with $\rho \approx 1.2\,\text{kg/m}^3$ the density of air. Take an illustrative restriction with $K = 9.0\,\text{Pa/(m}^3\text{/s)}^2$, known to $\pm 15\,\%$ from a commissioning test, and a reading $\Delta p = 22.5 \pm 0.5\,\text{Pa}$. Then $\dot{V} = \sqrt{22.5/9.0} = 1.58\,\text{m}^3\text{/s}$ and $\dot{m} = 1.2 \times 1.58 = 1.90\,\text{kg/s}$. The square root halves the relative uncertainties, so $\sigma_{\dot{m}}/\dot{m} = \tfrac{1}{2}\sqrt{(0.5/22.5)^2 + 0.15^2} = 0.076$, and the pressure route gives $1.90 \pm 0.14\,\text{kg/s}$.

The two estimates rest on different sensors and different physics, so they are independent evidence about one quantity. Their difference, $2.0 - 1.90 = 0.10\,\text{kg/s}$, has a standard deviation

of $\sqrt{0.15^2 + 0.14^2} = 0.21\,\mathrm{kg/s}$, and is half a standard deviation from zero: the two agree. Agreement raises confidence, and the estimates combine by inverse-variance weighting, with weights $1/0.15^2 = 44.4$ and $1/0.144^2 = 48.2$ (in $(\mathrm{kg/s})^{-2}$), into

$$\hat{\dot{m}} = \frac{44.4 \times 2.0 + 48.2 \times 1.90}{44.4 + 48.2} = 1.95\,\mathrm{kg/s}, \qquad \sigma = \frac{1}{\sqrt{92.6}} = 0.10\,\mathrm{kg/s},$$

narrower than either input. If instead the pressure route had given 2.6 kg/s with the same width, the difference of 0.6 kg/s would be $0.6/0.21 = 2.9$ standard deviations from zero, an event with a probability of about 0.004 if both estimates were correct. Disagreement of that size says that something is wrong: a biased exhaust sensor, a changed flow coefficient K from a loaded filter, or heat leaving the rack by a path the energy balance omits. Averaging the two would hide that statement.

Figure 14.1 draws both cases. The horizontal axis is the rack airflow in kilograms per second; the vertical axis is probability density. In panel (a) the solid red curve is the energy-balance estimate, the dashed blue curve the pressure estimate and the thick green curve their combination; in panel (b) the pressure estimate is moved to 2.6 kg/s. In panel (a) the two input curves overlap over most of their width and the combined curve is taller and narrower than either. In panel (b) they barely overlap, and no single airflow is plausible under both; the combination is not drawn because it would describe neither.

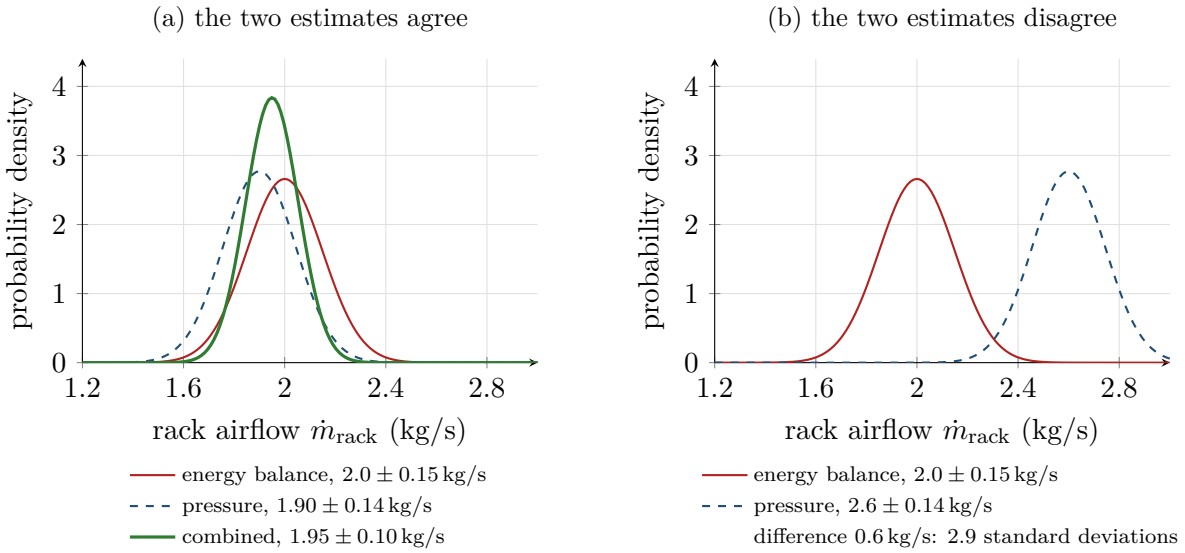


Figure 14.1: Two independent estimates of one rack airflow, drawn as normal densities (illustrative values). (a) The energy-balance estimate of equation (14.3) and the pressure estimate of equation (14.4) agree, and their inverse-variance combination is narrower than either. (b) The pressure estimate is moved to 2.6 kg/s; the estimates are inconsistent and are not combined.

14.1.6 The state vector of a cooling zone

For one cooling zone with n racks, a useful state vector is

$$x = \left[\, T_{\mathrm{cold}},\ T_{\mathrm{hot}},\ T_{\mathrm{rack},1}, \ldots, T_{\mathrm{rack},n},\ \dot{m}_{\mathrm{rack},1}, \ldots, \dot{m}_{\mathrm{rack},n},\ \dot{m}_w,\ p_{\mathrm{plenum}} \,\right], \tag{14.5}$$

where T_{cold} and T_{hot} are the cold-aisle and hot-aisle air temperatures, $T_{\mathrm{rack},i}$ and $\dot{m}_{\mathrm{rack},i}$ the temperature and airflow of rack i, $\dot{m}_w$ the chilled-water flow through the zone's coil and p_{plenum} the underfloor plenum pressure. The parameters are

$$\theta = \left[\, R_{\mathrm{th}},\ C_{\mathrm{th}},\ K_{\mathrm{air}},\ UA_{\mathrm{coil}},\ \alpha \,\right], \tag{14.6}$$

the thermal resistance and capacitance of the zone, the flow-resistance coefficient of its air path, the conductance of its coil and its recirculation fraction. The measurement vector is

$$y = \left[\text{temperature sensors},\ P_{\text{rack}},\ \Delta p,\ T_{\text{CHW,s}},\ T_{\text{CHW,r}},\ \text{flow meters}\right]. \tag{14.7}$$

For ten racks the state has 24 components and the parameters five, while a typical instrumentation set gives perhaps twenty to thirty readings, some of which, such as two temperature points in one aisle, carry almost the same information.

Figure 14.2 draws the three vectors and the relations between them. The box at the upper left is the state, the box at the lower left the parameters, the box in the middle the measurement model of equation (14.1), and the box on the right the measurements. Solid arrows run in the physical direction, from state and parameters through the sensors to the readings; the dashed arrows run in the direction of inference, from the readings through the estimator back to the state and parameters. The figure shows the asymmetry of the problem: the physics maps the left side to the right uniquely, but the right side has fewer independent entries than the left, so the reverse map needs the model's structure and a statement of uncertainty.

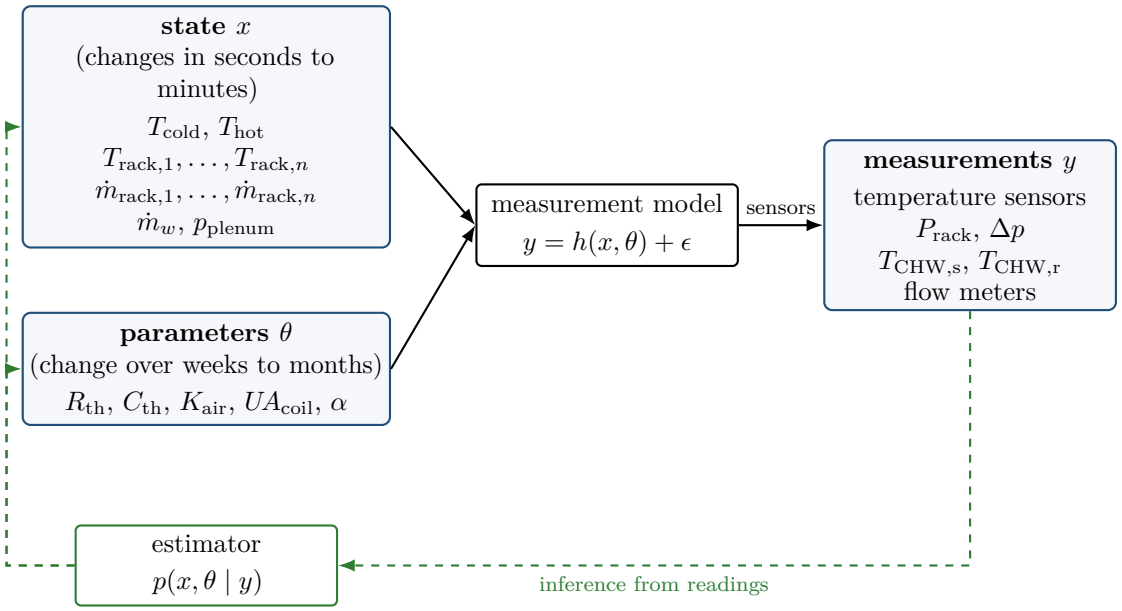


Figure 14.2: The state, parameter and measurement vectors of one cooling zone, equations (14.5) to (14.7). Solid arrows: the physical map from state and parameters to readings. Dashed arrows: the inference from readings back to a posterior over state and parameters.

14.1.7 State and parameter

The split between x and θ is a split by time scale. States change quickly: a temperature $T(t)$ responds within minutes to a load change, an airflow $\dot{m}(t)$ within seconds to a fan command, and the state of charge $\text{SOC}(t)$ of an uninterruptible power supply's battery within minutes of a discharge. Parameters change slowly: the thermal resistance R_{th} of a server's heat sink, the conductance UA of a coil and the flow coefficient K_{filter} of a filter are constant over a day. They are not constant over a year. A coil fouls, a filter loads with dust and a heat sink's thermal interface material dries out, so parameters degrade, and their degradation is often the most valuable thing to detect.

When a parameter must be tracked rather than fixed, it is appended to the state,

$$x_{\text{aug}} = \begin{bmatrix} x \\ \theta \end{bmatrix}, \qquad \theta_{k+1} = \theta_k + w_{\theta,k}, \tag{14.8}$$

with a random-walk model whose small noise $w_{\theta,k}$ allows the parameter to drift. The same estimator then follows the fast states and the slow parameters together. The size of $w_{\theta,k}$ encodes the time-scale split: a variance so small that the parameter barely moves in a day, but large enough that a decline over months is followed. Too large a value lets the parameter absorb every short disturbance; too small a value makes the estimator ignore a real degradation.

14.2 Dynamic state estimation

This section treats estimation as a process that runs in time: a model predicts the state forward, each new reading corrects the prediction, and the balance between the two is set by their uncertainties. It derives the Kalman filter for the linear case, works one update with numbers, and lists the methods for nonlinear models.

14.2.1 Process and measurement models

Dynamic estimation uses two models. The process model describes how the state evolves from one sample to the next,

$$x_{k+1} = f(x_k, u_k, d_k, \theta) + w_k, \tag{14.9}$$

where k counts samples, u_k are the controls (fan speeds, valve commands, setpoints), d_k the disturbances (IT load, outdoor conditions), f the discretized physics, for example the resistance–capacitance model of Chapter 8, and w_k the process noise: everything the model leaves out, from unmodeled airflow paths to load fluctuations below the sampling interval. The measurement model describes what the sensors report at each sample,

$$y_k = h(x_k) + v_k, \tag{14.10}$$

with v_k the measurement noise. Both noises are taken as zero-mean, independent from sample to sample and independent of each other, with covariances $\mathbf{\Sigma}_w$ and $\mathbf{\Sigma}_v$.

The two models are two sources of information about the same state. The process model says where the state should be now, given where it was and what was done to it; the measurement model says what the sensors should read if the state is where it is believed to be. An estimator uses both: what the physics predicts and what the sensors report. Neither is sufficient alone. A model run forward without measurements drifts away from reality as its errors accumulate; a measurement used without a model is as noisy as the sensor and says nothing about the states the sensors do not reach.

14.2.2 Prediction and correction

Each sample is processed in two steps. The prediction step propagates the previous estimate through the process model,

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_{k-1}, d_{k-1}), \tag{14.11}$$

where the subscript $k|k-1$ reads "the estimate at sample k using measurements up to sample $k-1$". The correction step compares the new reading with the reading the prediction implies. The difference,

$$r_k = y_k - h(\hat{x}_{k|k-1}), \tag{14.12}$$

is the innovation: the part of the new reading that the model did not expect. The corrected estimate $\hat{x}_{k|k}$ moves the prediction toward the reading by an amount proportional to r_k. Prediction then starts again from $\hat{x}_{k|k}$. This alternation is the idea behind Kalman filtering and, more generally, Bayesian filtering, in which the prediction step propagates the whole distribution

$p(x_{k-1} \mid y_{1:k-1})$ forward through the process model and the correction step multiplies it by the likelihood of the new reading, as in equation (14.2).

Figure 14.3 draws the loop. The boxes along the top are the prediction: the previous estimate and its covariance enter the process model, which also receives the controls and disturbances and adds the process noise covariance; the predicted state passes through the measurement model. The boxes along the bottom are the correction: the predicted reading is subtracted from the sensor reading, the innovation is weighted by the gain, and the result becomes the previous estimate of the next sample. The gain depends on the predicted covariance and on the measurement noise covariance, so the loop carries an uncertainty alongside every estimate.

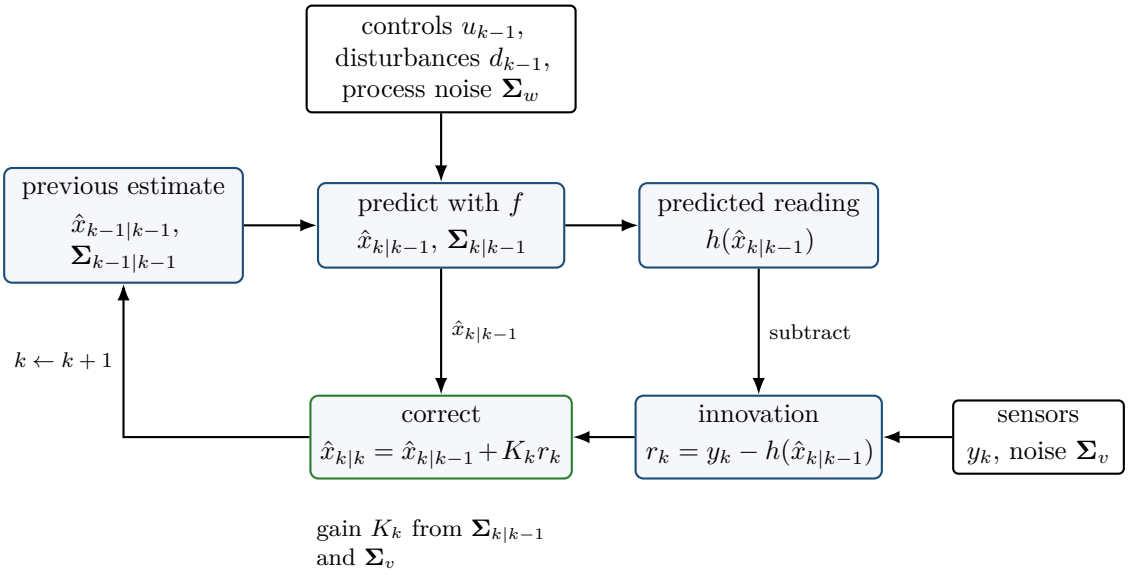


Figure 14.3: The prediction–correction loop of a recursive state estimator. Top row: prediction through the process model of equation (14.9) and the measurement model of equation (14.10). Bottom row: the innovation of equation (14.12) and the correction by the gain K_k.

14.2.3 The model as a plausibility check

The prediction does more than reduce noise; it states which readings are physically plausible. Take a rack inlet temperature whose previous estimate is 24.0°C with a standard deviation of 0.2 K, and a model that, given a known load increase nearby, predicts 24.3°C for the next minute with a standard deviation of 0.3 K. The sensor's noise has a standard deviation of 0.4 K, so the reading is expected to fall within a band of standard deviation $\sqrt{0.3^2 + 0.4^2} = 0.5$ K around 24.3°C.

The sensor then reports 30°C. The innovation is 5.7 K, or $5.7/0.5 = 11.4$ standard deviations, and the neighboring sensors report normal values. Two explanations compete: the inlet air has warmed by 6 K in one minute while the surrounding air has not, or the sensor has failed. The first requires a heat input that the measured rack and neighbor powers do not show and a rise that no thermal capacitance in the room allows in one minute without a matching rise nearby; the second requires only a loose connection or a failed transmitter. The sensor problem is far more plausible than a real 6 K jump. The dynamic model is what makes that judgment possible: it constrains what is physically plausible, and a reading far outside that constraint is evidence about the sensor, not about the state. A filter that applies this test rejects such a reading, or gives it a small weight, instead of following it; Section 14.3.4 develops the same test for a snapshot.

14.2.4 The Kalman filter

When the process and measurement models are linear,

$$x_{k+1} = \mathbf{A}x_k + \mathbf{B}u_k + w_k, \qquad y_k = \mathbf{H}x_k + v_k, \tag{14.13}$$

with matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{H}$, and the noises are normal, the posterior of the state stays normal at every step and is described by its mean $\hat{x}$ and covariance $\mathbf{\Sigma}$. The Kalman filter propagates both. The prediction step is

$$\hat{x}_{k|k-1} = \mathbf{A}\hat{x}_{k-1|k-1} + \mathbf{B}u_{k-1}, \qquad \mathbf{\Sigma}_{k|k-1} = \mathbf{A}\mathbf{\Sigma}_{k-1|k-1}\mathbf{A}^{\mathsf{T}} + \mathbf{\Sigma}_w, \tag{14.14}$$

and the correction step is

$$\begin{aligned} K_k &= \mathbf{\Sigma}_{k|k-1}\mathbf{H}^{\mathsf{T}}\left(\mathbf{H}\mathbf{\Sigma}_{k|k-1}\mathbf{H}^{\mathsf{T}} + \mathbf{\Sigma}_v\right)^{-1}, \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\left(y_k - \mathbf{H}\hat{x}_{k|k-1}\right), \\ \mathbf{\Sigma}_{k|k} &= (\mathbf{I} - K_k\mathbf{H})\mathbf{\Sigma}_{k|k-1}, \end{aligned} \tag{14.15}$$

with $\mathbf{I}$ the identity matrix. The new estimate is the model's prediction plus a weighted residual. The covariance grows in the prediction, by the process noise, and shrinks in the correction, by the information in the reading.

For one state measured directly, $\mathbf{H} = 1$, and the gain reduces to

$$K_k = \frac{\sigma^2_{k|k-1}}{\sigma^2_{k|k-1} + \sigma^2_v}, \qquad \sigma^2_{k|k} = (1 - K_k)\,\sigma^2_{k|k-1}, \tag{14.16}$$

where $\sigma^2_{k|k-1}$ is the variance of the prediction and σ^2_v that of the sensor. With the numbers of Section 14.2.3: the previous variance is $0.2^2 = 0.04\,\mathrm{K}^2$ and the process noise adds $0.05\,\mathrm{K}^2$ over the minute, so $\sigma^2_{k|k-1} = 0.09\,\mathrm{K}^2$ ($0.3\,\mathrm{K}$); the sensor variance is $0.4^2 = 0.16\,\mathrm{K}^2$. The gain is $K_k = 0.09/(0.09 + 0.16) = 0.36$. If the sensor reports a plausible 24.9°C, the innovation is $0.6\,\mathrm{K}$ and

$$\hat{x}_{k|k} = 24.3 + 0.36 \times 0.6 = 24.52^{\circ}\mathrm{C}, \qquad \sigma^2_{k|k} = (1 - 0.36) \times 0.09 = 0.058\,\mathrm{K}^2 \ (0.24\,\mathrm{K}).$$

The estimate moves a little more than a third of the way from the prediction to the reading, and its standard deviation, $0.24\,\mathrm{K}$, is smaller than both the prediction's $0.3\,\mathrm{K}$ and the sensor's $0.4\,\mathrm{K}$.

Figure 14.4 runs the scalar filter on a synthetic two-hour record, one reading per minute. The horizontal axis is time in minutes, the vertical axis the inlet temperature. The dashed black line is the true temperature, which rises from 24°C toward 26°C after a known load change at minute 40; gray dots are the readings, with a noise of $0.4\,\mathrm{K}$; the blue line is the filter estimate and the shaded band its $\pm 2\sigma$ interval. The filter starts at a deliberately wrong 23°C with a large variance, so its first gain is 0.86 and it moves almost all the way to the first reading; within a few minutes it settles to a steady gain of 0.16. After the load change it follows the rise because the model predicts it, not because the readings force it. At minute 80 the sensor reports 30.0°C, a glitch whose innovation is 9.2 standard deviations; the filter rejects readings beyond four standard deviations and keeps its prediction. An ungated filter would have been pulled to 26.64°C by that reading, $0.7\,\mathrm{K}$ above the truth. Over minutes 20 to 119 the root-mean-square error is $0.40\,\mathrm{K}$ for the raw readings and $0.13\,\mathrm{K}$ for the filter.

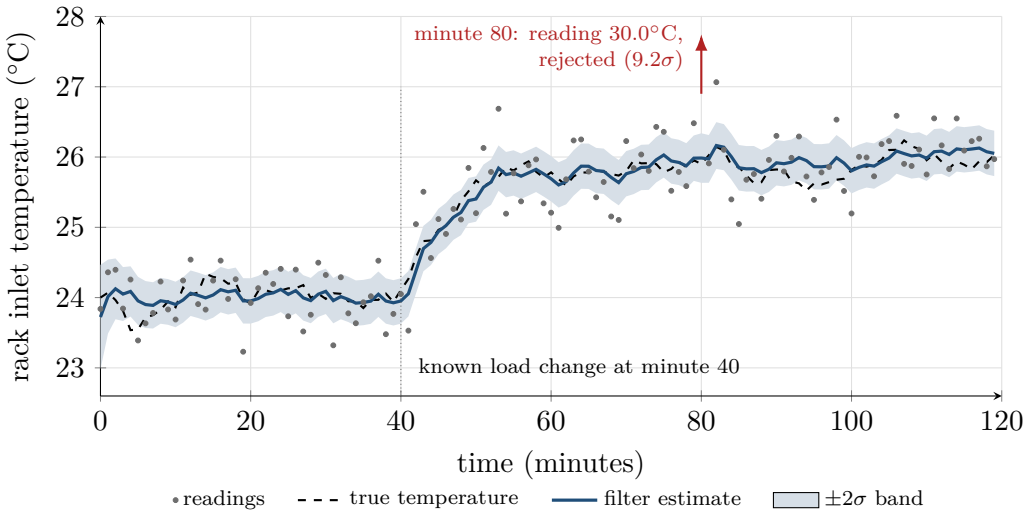


Figure 14.4: A scalar Kalman filter on a synthetic rack inlet temperature (illustrative, seed 14). Process model $x_{k+1} = x_k + 0.1\,(u_k - x_k) + w_k$ with $u_k = 24$°C before minute 40 and 26°C after, $\sigma_w = 0.1$ K; readings $y_k = x_k + v_k$ with $\sigma_v = 0.4$ K; filter started at 23°C with $\sigma = 1$ K; readings beyond four standard deviations of the innovation are rejected. Data from `scripts/ch14_data.py`.

14.2.5 The weighting between model and sensor

Equation (14.16) makes the weighting explicit. Writing $\gamma = \sigma^2_{k|k-1}/\sigma^2_v$ for the ratio of the prediction's variance to the sensor's,

$$K_k = \frac{\gamma}{1+\gamma}, \qquad \frac{\sigma_{k|k}}{\sigma_{k|k-1}} = \sqrt{1-K_k} = \frac{1}{\sqrt{1+\gamma}}. \tag{14.17}$$

An accurate sensor (small σ^2_v, large γ) gives a gain near one: the estimate follows the measurement. An accurate model (small process noise and therefore small $\sigma^2_{k|k-1}$, small γ) gives a gain near zero: the estimate follows the prediction and the reading barely moves it. The gain is the quantitative form of a judgment that operators make informally, how far to trust the physics against how far to trust the sensors, and it is set by the two noise covariances. Choosing them is a modeling decision: a process noise that is too small makes the filter slow to follow real changes, and one that is too large makes it as noisy as the sensor.

Figure 14.5 plots equation (14.17). The horizontal axis is the variance ratio γ on a logarithmic scale; the vertical axis is dimensionless. The solid curve is the gain; the dashed curve is the ratio of the corrected to the predicted standard deviation, which says how much one reading narrows the estimate. The two marked points are the worked update of Section 14.2.4, $\gamma = 0.09/0.16 = 0.56$ and $K = 0.36$, and the steady state of the filter in Figure 14.4, $\gamma = 0.176^2/0.16 = 0.19$ and $K = 0.16$. Across four decades of γ the gain moves from model to sensor along an S-shaped curve, and one reading reduces the standard deviation substantially only when γ is near one or larger.

14.2.6 Nonlinear estimation

Much of a data center's physics is nonlinear. The pressure across a restriction grows with the square of the flow, $\Delta p = K\dot{V}^2$; the power of a fan grows with the cube of its speed, $P_{\text{fan}} \propto N^3_{\text{fan}}$ (Chapter 6); the heat a coil removes is a product of a flow and a temperature difference. The Kalman filter's equations do not apply directly, and several methods extend them. The *extended Kalman filter* linearizes f and h around the current estimate and applies equation (14.15) with

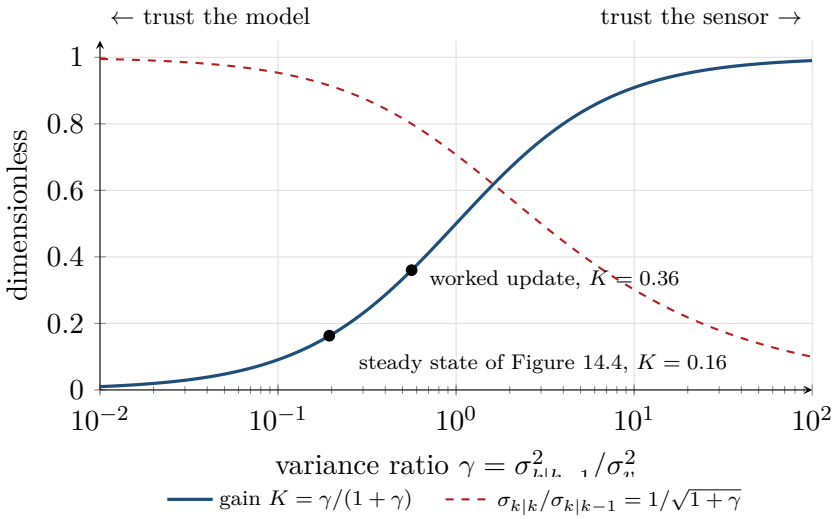


Figure 14.5: The scalar Kalman gain and the reduction of the standard deviation by one reading, as functions of the ratio of prediction variance to sensor variance, equation (14.17). Markers: the worked update of Section 14.2.4 and the steady state of the filter in Figure 14.4.

the Jacobian matrices in place of $\mathbf{A}$ and $\mathbf{H}$. For the pressure relation at the illustrative operating point of Section 14.1.5, $\dot{V}_0 = 1.58\,\mathrm{m}^3/\mathrm{s}$, the linearization is

$$\frac{\partial \Delta p}{\partial \dot{V}} = 2K\dot{V}_0 = 2 \times 9.0 \times 1.58 = 28.5\,\mathrm{Pa}/(\mathrm{m}^3/\mathrm{s}),$$

so a pressure sensor with a 0.5 Pa error constrains the flow to about $0.5/28.5 = 0.018\,\mathrm{m}^3/\mathrm{s}$ near that point, if K is known; at half the flow the slope halves and the same sensor constrains the flow half as well.

The *unscented Kalman filter* propagates a small set of chosen sample points through the nonlinear functions instead of linearizing them, which handles stronger curvature. A *particle filter* represents the posterior by a large set of weighted samples and can follow a distribution of any shape, including one with two peaks, at the cost of many model evaluations per step. *Nonlinear optimization* over a window of samples (moving-horizon estimation) solves for the state trajectory that best fits the model and the data together, and handles bounds such as a valve position between zero and one. Full *Bayesian inference* by sampling computes the posterior of states and parameters jointly and is the reference against which the approximations are judged. The choice among them matters less than the formulation they share: physics plus measurements, each with a stated uncertainty, give a state estimate with a stated uncertainty. A correct formulation solved by a simple method is more useful than a wrong one solved by an elaborate method.

14.3 Static estimation and residuals

This section treats estimation from a single snapshot of readings, without a process model: the weighted least-squares estimator, its counterpart in power systems, the residuals it leaves behind, and how those residuals expose a bad measurement.

14.3.1 Weighted least squares

A static estimator uses one set of simultaneous readings and the algebraic relations among them. It chooses the state that makes the predicted readings closest to the actual ones, with

each reading weighted by its accuracy,

$$\hat{x} = \arg\min_{x} \; (y - h(x))^{\mathsf{T}} \mathbf{W} (y - h(x)), \qquad \mathbf{W} = \boldsymbol{\Sigma}_v^{-1}, \tag{14.18}$$

where $\mathbf{W}$ is the inverse of the measurement covariance, so that a reading with half the standard deviation counts four times as much. For a linear model $h(x) = \mathbf{H}x$ the minimum has the closed form

$$\hat{x} = (\mathbf{H}^{\mathsf{T}}\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^{\mathsf{T}}\mathbf{W}\, y, \qquad \operatorname{Cov}(\hat{x}) = (\mathbf{H}^{\mathsf{T}}\mathbf{W}\mathbf{H})^{-1}. \tag{14.19}$$

With normal errors and no prior information, $\hat{x}$ is the maximum of the likelihood and the second expression is its covariance. The inverse-variance combination of Section 14.1.5 is the special case of one state measured twice.

A small example shows the structure. Two racks are fed by one cooling unit. The state is the pair of rack airflows, $x = [\dot{m}_1, \dot{m}_2]$. Three readings bear on it: the cooling unit's total airflow from its fan curve, $y_1 = 4.1 \pm 0.2\,\mathrm{kg/s}$, which measures $\dot{m}_1 + \dot{m}_2$; and the energy-balance estimates of the two racks, $y_2 = 2.0 \pm 0.15\,\mathrm{kg/s}$ and $y_3 = 1.9 \pm 0.15\,\mathrm{kg/s}$, which measure each airflow alone. Then

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{W} = \operatorname{diag}(25,\ 44.4,\ 44.4), \quad \mathbf{H}^{\mathsf{T}}\mathbf{W}\mathbf{H} = \begin{bmatrix} 69.4 & 25 \\ 25 & 69.4 \end{bmatrix}, \quad \mathbf{H}^{\mathsf{T}}\mathbf{W}y = \begin{bmatrix} 191.4 \\ 186.9 \end{bmatrix},$$

with weights in $(\mathrm{kg/s})^{-2}$. Solving gives $\hat{\dot{m}}_1 = 2.053\,\mathrm{kg/s}$ and $\hat{\dot{m}}_2 = 1.953\,\mathrm{kg/s}$, each with a standard deviation of $0.129\,\mathrm{kg/s}$ and a correlation of -0.36 between them. Table 14.1 lists the readings, the predicted readings $h(\hat{x})$ and the residuals. Each rack estimate is tighter than its own energy balance, 0.129 against $0.15\,\mathrm{kg/s}$, because the total airflow adds information; the negative correlation says that the total pins the sum more tightly than the split.

Table 14.1: Weighted least-squares estimate of two rack airflows from three readings (illustrative). The normalized residual is the residual divided by the reading's standard deviation.

Reading	Measures	y (kg/s)	σ (kg/s)	$h(\hat{x})$ (kg/s)	r/σ
cooling-unit fan curve	$\dot{m}_1 + \dot{m}_2$	4.10	0.20	4.006	+0.47
rack 1 energy balance	$\dot{m}_1$	2.00	0.15	2.053	−0.35
rack 2 energy balance	$\dot{m}_2$	1.90	0.15	1.953	−0.35

14.3.2 The analogy with power-system state estimation

Weighted least squares is the standard method of state estimation in transmission systems, and the analogy is close. A transmission operator's estimator computes, every few minutes, the voltage magnitude $|V_i|$ and phase angle δ_i at every bus i of the network (written δ_i here, to keep θ for parameters). Its measurements are power injections at buses, power flows on lines and voltage magnitudes, each related to the bus voltages by the nonlinear power-flow equations, and it solves equation (14.18) by repeated linearization. The measurements are redundant, typically two to three times as many as the states, and the redundancy is what lets the estimator detect a bad meter or a wrong breaker status.

A data center's thermal and airflow network has the same structure. The hidden quantities are the airflows $\dot{m}_i$, temperatures T_i, heat flows $\dot{Q}_i$ and pressures p_i at the nodes and edges of the air and water paths; the constraints are the conservation of mass and energy at each node and the flow–pressure relations on each edge; the measurements are temperatures, pressures, powers and a few flows. Two differences matter. The electrical network is metered densely and its equations are known accurately, while the thermal network has sparse flow measurements and

uncertain parameters such as recirculation fractions. And the electrical estimator's states are nearly all algebraic at its time scale, while the thermal network's temperatures have capacitance and respond over minutes, which is why the dynamic estimators of Section 14.2 are more often needed. Thermal state estimation is nonetheless the same problem as electrical state estimation: hidden node and edge quantities constrained by conservation laws and observed through a redundant, imperfect set of meters.

14.3.3 Residuals

Once $\hat{x}$ is computed, it predicts every reading, $\hat{y} = h(\hat{x})$, and the residuals

$$r = y - \hat{y} = y - h(\hat{x}) \tag{14.20}$$

say how well the estimated state explains the data. With correct models and normal errors, the normalized residuals r_j/σ_j are mostly within ± 2 and the weighted sum of squares $\chi^2 = \sum_j r_j^2/\sigma_j^2$ follows a chi-squared distribution with $m - n$ degrees of freedom, where m is the number of readings and n the number of states. In the example of Table 14.1, $m - n = 3 - 2 = 1$ and $\chi^2 = 0.47^2 + 0.35^2 + 0.35^2 = 0.47$, well below the 99th percentile of that distribution, 6.6: the three readings are consistent.

Large residuals have four kinds of cause. A bad measurement: a failed, biased or mislabeled sensor. A model error: a wrong parameter or a missing term, such as an omitted heat path. A topology error: the model connects a rack to the wrong cooling unit, or a damper assumed open is closed. An unmodeled fault: a blocked filter or a stuck valve that the model does not represent. State estimation and bad-data detection therefore belong together: the estimate supplies the predictions that make the residuals, and the residuals test the estimate. With $m = n$, every reading is needed to fix the state and every residual is zero whatever the errors, so the test needs redundancy, $m > n$.

14.3.4 A bad temperature measurement

Five nearby inlet sensors report 23.9, 24.1, 24.0, 24.2 and 31.5°C; the model, using the measured supply temperature and rack powers, predicts about 24.0°C at all five, and each sensor has a standard deviation of 0.3 K. The residuals against the model are −0.1, +0.1, 0.0, +0.2 and +7.5 K. The first four are within one standard deviation; the fifth is $7.5/0.3 = 25$ standard deviations from zero, and the weighted sum of squares is dominated by it. The fifth reading is bad data.

A static estimate that includes the bad reading spreads its error over all five. If the five readings are used to estimate one common inlet temperature by equation (14.18), the estimate is their mean, $127.7/5 = 25.54$°C, 1.5 K above the 24.05°C that the four good sensors give. The residuals against that estimate are −1.64, −1.44, −1.54, −1.34 and +5.96 K: the bad reading's residual shrinks and the good readings' residuals grow to about five standard deviations each. This smearing is why bad data are removed one at a time, starting with the largest normalized residual, and the estimate is recomputed after each removal.

Figure 14.6 draws both sets of residuals. The horizontal axis lists the five sensors with their readings; the vertical axis is the residual in kelvin. Dark bars are residuals against the model prediction of 24.0°C; light bars are residuals against the common estimate that includes the bad reading; the dashed lines mark $\pm 3\sigma = \pm 0.9$ K. Against the model only sensor 5 leaves the band. Against the contaminated estimate every sensor leaves it, and the culprit is less obvious.

The residual says that the fifth reading is inconsistent with the model and the other readings. It does not say why. Three explanations remain: the sensor has failed or drifted; there is a real local hot spot at that point, such as exhaust escaping through a missing blanking panel, which the model does not represent; or the sensor is mapped to the wrong location and is

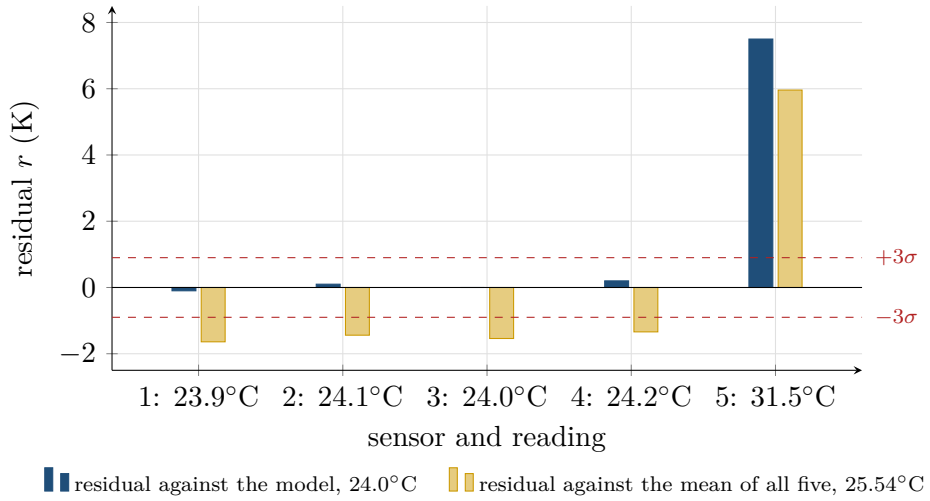


Figure 14.6: Residuals of five inlet sensors (illustrative), with a sensor standard deviation of 0.3 K. Against the model prediction only sensor 5 is outside $\pm 3\sigma$; against an estimate that includes sensor 5, its error is spread over all five.

in fact reading a hot aisle. The three call for different actions, replacing a sensor, fixing the airflow or correcting a database, and separating them is the task of diagnosis (Chapter 13 and Chapter 16), which needs evidence beyond the one residual: a second sensor at the same point, the history of the reading, the neighbor rack's exhaust temperature.

14.4 Estimated quantities in a data center

This section applies estimation to five hidden quantities that matter for operation: the recirculation fraction at a rack inlet, the conductance of a cooling coil, the actual position of a valve, the available cooling capacity of a plant, and the topology that connects cooling units to racks. Each is inferred, not measured, and each is reported with an uncertainty.

14.4.1 Recirculation fraction

The mixing rule of Chapter 6 writes a rack's inlet temperature as a blend of supply air and hot exhaust, $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{hot}}$, which gives

$$\alpha = \frac{T_{\text{in}} - T_{\text{supply}}}{T_{\text{hot}} - T_{\text{supply}}}. \tag{14.21}$$

With illustrative readings $T_{\text{supply}} = 20°\text{C}$, $T_{\text{hot}} = 40°\text{C}$ and $T_{\text{in}} = 25°\text{C}$, $\alpha = 5/20 = 0.25$: a quarter of the air entering the rack is recirculated exhaust.

The estimate is sensitive to its three inputs in different degrees. The partial derivatives at this point are $\partial\alpha/\partial T_{\text{in}} = 1/20 = 0.050\,\text{K}^{-1}$, $\partial\alpha/\partial T_{\text{supply}} = -15/20^2 = -0.0375\,\text{K}^{-1}$ and $\partial\alpha/\partial T_{\text{hot}} = -5/20^2 = -0.0125\,\text{K}^{-1}$. With ± 0.5 K on each sensor, the standard deviation is

$$\sigma_\alpha = 0.5\sqrt{0.050^2 + 0.0375^2 + 0.0125^2} = 0.032,$$

so $\alpha = 0.25 \pm 0.03$. The inlet sensor matters most and the hot-aisle sensor least. The representativeness of T_{hot} is the larger issue: the mixing rule needs the temperature of the exhaust that actually reaches the inlet, which is not the hot-aisle average, so α estimated this way is an effective fraction for the chosen sensor, not a measured property of the airflow.

14.4.2 Coil conductance and its degradation

The heat a cooling coil transfers is related to its overall conductance by $\dot{Q} \approx UA\,\Delta T_{\mathrm{lm}}$ (Chapter 7), where the log-mean temperature difference for a counterflow coil with end differences ΔT_1 and ΔT_2 is

$$\Delta T_{\mathrm{lm}} = \frac{\Delta T_1 - \Delta T_2}{\ln(\Delta T_1/\Delta T_2)}, \qquad \widehat{UA} = \frac{\dot{Q}}{\Delta T_{\mathrm{lm}}}. \tag{14.22}$$

With the heat flow measured on the water side and the four temperatures measured, UA is estimated. For an illustrative coil, air enters at 32°C and leaves at 20°C, and $\dot{m}_w = 12\,\mathrm{kg/s}$ of water enters at 14°C and leaves at 20°C. The heat flow is $\dot{Q} = 12 \times 4.18 \times 6 = 301\,\mathrm{kW}$; the end differences are $\Delta T_1 = 32 - 20 = 12\,\mathrm{K}$ and $\Delta T_2 = 20 - 14 = 6\,\mathrm{K}$; so $\Delta T_{\mathrm{lm}} = 6/\ln 2 = 8.66\,\mathrm{K}$ and $\widehat{UA} = 301/8.66 = 34.8\,\mathrm{kW/K}$.

A single estimate says little. A sequence of estimates over months, each computed at comparable operating conditions, says whether the coil is degrading. Table 14.2 lists an illustrative sequence. Each estimate carries about ±5 % from its sensors, so a change between consecutive estimates is within noise, but the trend over ten months, a 13 % decline, is not; and a bias that is the same in every month cancels from the trend. The last column shows the consequence: to transfer the same 301 kW the coil needs a larger log-mean difference, which means colder water, more water or warmer return air. A falling $UA(t)$ indicates fouling of the coil surfaces, degraded heat transfer on the water side, or an airflow problem across the coil face. Tracking it is condition estimation: the parameter θ of equation (14.8) followed as a slowly drifting state.

Table 14.2: Monthly estimates of a coil's conductance at a duty of 301 kW (illustrative). Each estimate is uncertain by about ±5 %; the trend over ten months is the signal.

Month	$\widehat{UA}$ (kW/K)	Relative to month 0 (%)	ΔT_{lm} needed for 301 kW (K)
0	34.8	100.0	8.65
2	34.6	99.4	8.70
4	33.9	97.4	8.88
6	33.0	94.8	9.12
8	31.8	91.4	9.46
10	30.4	87.4	9.90

14.4.3 Valve state

A chilled-water valve receives a command u_{cmd} and may report a position feedback, but its actual opening u_{actual} is a hidden state (Chapter 9). It can be inferred from the flow it passes. Take an illustrative valve that passes 15 kg/s fully open at the loop's normal differential pressure and has a linear characteristic, so that the flow is proportional to the opening. The command is $u_{\mathrm{cmd}} = 80\,\%$, which should pass 12 kg/s; the flow meter reads 4.5 kg/s, which corresponds to

$$u_{\mathrm{actual}} \approx \frac{4.5}{15} = 30\,\%.$$

The commanded and inferred positions differ by 50 percentage points, far beyond any flow-meter error.

The discrepancy has several explanations. The valve may be stuck at about 30 %. The actuator's position feedback may report 80 % while the stem has not moved, so the control system believes the command was executed. The valve may be open as commanded while a strainer or a partly closed manual valve upstream obstructs the flow, or the differential pressure across the valve may have fallen. Each produces the same low flow at the same command.

What the estimate establishes is that the hydraulic state is not the commanded state; which explanation holds is again a diagnosis, which a differential-pressure reading across the valve or a stroke test would resolve.

14.4.4 Available cooling capacity

The cooling a plant can deliver now, $\dot{Q}_{\text{avail}}$, is not its nameplate rating. It depends on the equipment in service, the water and air flows, the chilled-water and condenser temperatures, the weather, the efficiency of the compressors and the degradation of every heat exchanger. It is therefore a state to be estimated in real time, from the same measurements and models as the other states, and reported with an uncertainty.

Take an illustrative chiller with a nameplate capacity of 4 MW at design conditions. On a hot afternoon three effects reduce what it can deliver: the high ambient temperature raises the condensing pressure, costing about 0.45 ± 0.20 MW; condenser fouling raises it further, costing about 0.25 ± 0.15 MW; and partial compressor degradation, seen as a lower flow at a given speed, costs about 0.20 ± 0.15 MW. With the three effects estimated independently,

$$\dot{Q}_{\text{avail}} = 4 - 0.45 - 0.25 - 0.20 = 3.1\,\text{MW}, \qquad \sigma = \sqrt{0.20^2 + 0.15^2 + 0.15^2} = 0.29\,\text{MW},$$

so $\dot{Q}_{\text{avail}} = 3.1 \pm 0.3$ MW. The nameplate overstates the capacity by almost 0.9 MW, and the uncertainty is a third of that gap. A capacity plan that uses 4 MW is wrong; one that uses 3.1 MW without its ± 0.3 MW cannot state how likely a shortfall is. Section 14.6.1 turns that uncertainty into a probability.

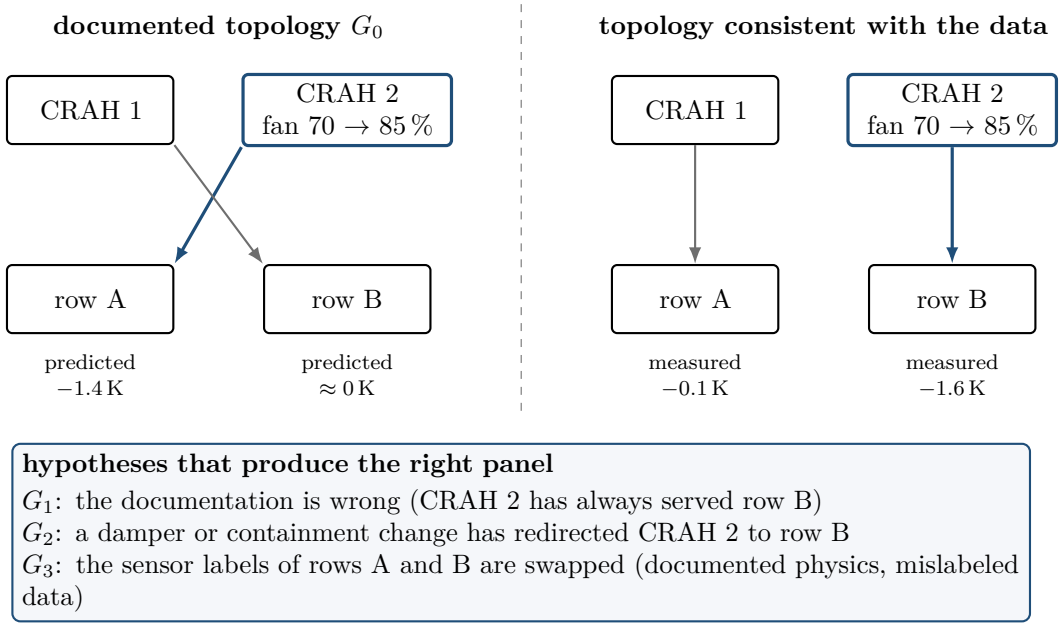
14.4.5 Uncertain topology

Most estimators assume that the topology, which cooling unit serves which racks, which pipe feeds which coil, which damper is open, is known and correct. In a real facility it is documented, and documentation lags reality: dampers are adjusted, containment panels moved, sensors relabeled during maintenance. When the topology G is uncertain, the object to estimate is the joint posterior of the continuous state and the discrete structure,

$$p(x, G \mid y) = p(x \mid y, G)\ \Pr(G \mid y), \tag{14.23}$$

which is harder than state estimation alone: for each candidate topology a separate state estimate is computed, and the candidates are weighted by how well each explains the data, a comparison of models rather than a fit within one.

An illustrative case shows the kind of evidence involved. The documentation says that cooling unit CRAH 2 serves row A and CRAH 1 serves row B. CRAH 2 raises its fan speed from 70 % to 85 %. Under the documented topology the model predicts that the row A inlet temperatures fall by about 1.4 K and the row B inlets barely change. The measurements show the reverse: row B falls by 1.6 K and row A by 0.1 K. The documented topology is inconsistent with the response. Figure 14.7 draws the situation. In the left panel, boxes are the two cooling units and the two rows and arrows are the airflow paths of the documented topology, with the predicted response under each row; in the right panel the arrows are the paths of the topology consistent with the data, with the measured response. The box below lists the hypotheses that produce the right panel. A swapped damper configuration and a wrong document predict the same physics; a swap of the two rows' sensor labels produces the same readings with the documented physics. The step test separates the documented topology from the rest, but not those three from each other; that takes a further measurement, such as a portable temperature reading at row A during a second step. This is structural inference: an inference about the graph, not about a value on it.



hypotheses that produce the right panel
G_1: the documentation is wrong (CRAH 2 has always served row B)
G_2: a damper or containment change has redirected CRAH 2 to row B
G_3: the sensor labels of rows A and B are swapped (documented physics, mislabeled data)

Figure 14.7: Topology inference from a fan-speed step (illustrative). Left: the documented topology and the response it predicts. Right: the topology consistent with the measured response. Thick blue arrows are the paths of the unit whose fan speed changed. The three hypotheses below explain the data equally well and need a further measurement to separate.

14.5 Observability and posterior uncertainty

This section states when the measurements determine the state at all, why more samples of the same kind do not help, how a transient adds information that a steady state lacks, and why an estimate must carry the shape of its posterior and not only its center.

14.5.1 Observability

The first question about any estimator comes before the estimator: do the measurements contain enough information to determine the states of interest? If two different states produce the same readings,

$$x_1 \neq x_2 \quad \text{and} \quad h(x_1) = h(x_2), \tag{14.24}$$

no estimator can tell them apart from those readings, however accurate the sensors and however clever the algorithm. The state is then unobservable under the current sensor set. For a linear static model $y = \mathbf{H}x$, the state is observable when $\mathbf{H}$ has as many independent rows as x has components, its rank equals n; any direction Δx with $\mathbf{H}\Delta x = 0$ can be added to the state without changing a reading. For a linear dynamic model the condition extends to the stacked matrix $[\mathbf{H};\ \mathbf{HA};\ \ldots;\ \mathbf{HA}^{n-1}]$, because a state that does not appear in the readings now may appear in them later, after the dynamics move it.

Observability is a structural property: it depends on which functions of the state are measured, not on their noise. It can therefore be checked before any data are collected, from the model and the sensor list, and it is the third step of the workflow in Section 14.6.6. A partial answer is common: some components of the state are observable and others are not, and the estimator should report the second group as unresolved rather than return a value for them.

14.5.2 An unobservable example

Suppose only the temperature rise across a rack is known, $T_{\text{out}} - T_{\text{in}} = 10\,\text{K}$, and neither its power nor its airflow is measured. The relation $\dot{Q} = \dot{m} c_{p,\text{air}} \Delta T$ becomes $\dot{Q} = 10\,\text{kJ/kg} \times \dot{m}$, a line in the plane of $\dot{m}$ and $\dot{Q}$: 1 kg/s with 10 kW, 2 kg/s with 20 kW, 3 kg/s with 30 kW all produce the same reading. Neither $\dot{Q}$ nor $\dot{m}$ is identifiable; only their ratio is. The Jacobian of $h(\dot{m}, \dot{Q}) = \dot{Q}/(\dot{m} c_{p,\text{air}})$ has one row and rank one, against two unknowns. This is a limitation of the physics and the data, not of an algorithm: no filter, optimizer or neural network recovers what the readings do not contain. A prior can still produce a posterior, but its spread along the line is the prior's spread; the data have narrowed only the direction across it. Metering the rack power, as in Section 14.1.3, adds the missing row and makes both quantities observable.

More samples are not more information. If the same rack reports $\Delta T = 10\,\text{K}$ every second for a year, the record holds about 3.2×10^7 samples. They determine ΔT very precisely, and they leave the airflow exactly as undetermined as one sample did, because every sample adds information about the same ratio and none about the direction along the line. More samples reduce noise in the observable directions; they do nothing for the unobservable ones. What helps is a sample of a different kind: a different function of the state, or the same function under different conditions.

14.5.3 Dynamic excitation

Different conditions are what a transient supplies. Model the rack's exhaust as one thermal node,

$$C_{\text{th}} \frac{\mathrm{d}T_{\text{out}}}{\mathrm{d}t} = P_{\text{rack}} - \dot{m} c_{p,\text{air}} \, (T_{\text{out}} - T_{\text{in}}), \tag{14.25}$$

with time constant $\tau = C_{\text{th}}/(\dot{m} c_{p,\text{air}})$, the resistance–capacitance form $\tau = R_{\text{th}} C_{\text{th}}$ of Chapter 8 with $R_{\text{th}} = 1/(\dot{m} c_{p,\text{air}})$. At steady state the left side is zero, and the readings determine only $\dot{m} c_{p,\text{air}} = P_{\text{rack}}/\Delta T$; the capacitance C_{th} does not appear and cannot be estimated from any amount of steady data. A change of power reveals it. If the rack power steps from 20 to 30 kW at an airflow of 2.0 kg/s, the exhaust rise moves from 10 to 15 K along an exponential with time constant τ; an observed $\tau = 120\,\text{s}$ gives $C_{\text{th}} = \tau \, \dot{m} c_{p,\text{air}} = 120 \times 2.0 \times 1000 = 2.4 \times 10^5\,\text{J/K}$.

When the power is not metered the case is stronger still. At steady state the readings give only $P_{\text{rack}}/\dot{m}$, as in Section 14.5.2; during a transient the shape of the response gives τ and the size of its change gives a ratio of the step to the airflow, so more combinations of C_{th}, R_{th} and $\dot{m}$ become separable. The same structure holds across the facility: quantities unidentifiable at steady state become identifiable in a transient. Natural transients (a load change, a unit switching off) are useful but uncontrolled. A deliberate test, a load-bank step, a setpoint step or a fan-speed step as in Section 14.4.5, is designed to excite exactly the direction that the steady data leave undetermined, which is why controlled testing is valuable. Chapter 15 develops identifiability, the parameter-side counterpart of observability, in full.

14.5.4 Posterior uncertainty

An estimator should return not only $\hat{x}$ but its uncertainty: an airflow of 2.1 ± 0.2 kg/s, or the full posterior $p(\dot{m} \mid y)$. The reason is that decisions downstream depend on how confident the estimate is, not only on its value. A rack whose airflow is 2.1 ± 0.2 kg/s against a requirement of 1.5 kg/s is adequately cooled with high confidence; the same center with ± 0.8 kg/s is not.

Two racks with the same estimate $\hat{\dot{m}} = 2.0$ kg/s make the point. For rack 1 the standard deviation is 0.05 kg/s: its airflow is well determined. For rack 2 it is 0.8 kg/s: its exhaust sensor is poorly placed, or its power is not metered. The probability that the airflow is below a

requirement of 1.5 kg/s is, with normal posteriors,

$$\text{rack 1:}\quad \Pr(\dot{m} < 1.5) = \Phi\left(\frac{1.5 - 2.0}{0.05}\right) = \Phi(-10) \approx 0,$$
$$\text{rack 2:}\quad \Pr(\dot{m} < 1.5) = \Phi\left(\frac{1.5 - 2.0}{0.8}\right) = \Phi(-0.63) = 0.27,$$

where Φ is the standard normal distribution function. The two racks must not be treated identically. Rack 1 needs nothing; rack 2 has a one-in-four chance of being underfed and needs either a better measurement or a conservative action. A report that lists both as 2.0 kg/s hides the only fact that distinguishes them.

14.5.5 The MAP estimate and a bimodal posterior

When one number must be reported, a common choice is the maximum a posteriori estimate,

$$x_{\text{MAP}} = \arg\max_x \; p(x \mid y), \tag{14.26}$$

the single most plausible state. With a normal posterior it equals the mean, and with a flat prior it equals the weighted least-squares estimate. It gives one best estimate and loses the rest of the distribution: its width, and anything else about its shape.

The shape matters when the data support two different states. Suppose the readings are consistent with two operating modes of a rack: fans at low speed with an airflow near 1.0 kg/s, or fans at high speed with an airflow near 2.5 kg/s, because the exhaust sensor is ambiguous between a low flow with little recirculation and a high flow with more. The posterior is bimodal, here an illustrative equal mixture of $\mathcal{N}(1.0, 0.15^2)$ and $\mathcal{N}(2.5, 0.20^2)$. Its mean is 1.75 kg/s, and its density at the mean is 0.0009, against 1.33 at the lower peak: the mean describes an airflow that is almost impossible. The MAP picks the taller, narrower peak at 1.0 kg/s and discards the other mode, which holds half the probability. Neither single number is an adequate report; the adequate report is "either about 1.0 kg/s or about 2.5 kg/s, with roughly equal probability," and the useful next step is a measurement that separates the two modes, such as the fan speed. Bimodal posteriors arise wherever competing operating modes or fault hypotheses explain the same data, which is the setting of probabilistic diagnosis.

Figure 14.8 draws both situations. In both panels the horizontal axis is the rack airflow and the vertical axis is probability density. In panel (a) the tall narrow curve is rack 1 and the low wide curve rack 2, with the area below 1.5 kg/s shaded for rack 2. In panel (b) the curve is the bimodal posterior, the dashed line its mean and the dotted line its MAP. Panel (a) shows two estimates that a table would report identically and that carry different risk; panel (b) shows that the mean can sit where the posterior has almost no mass.

14.6 Use and limits of state estimation

This section follows the estimate downstream, into risk, control, diagnosis and forecasting; separates real-time filtering from after-the-fact smoothing; states the ways an estimate fails when the model is wrong; and collects the procedure into a workflow.

14.6.1 From state uncertainty to risk

An uncertain state makes every quantity computed from it uncertain. The cooling margin of a plant,

$$M = \dot{Q}_{\text{avail}} - \dot{Q}_{\text{req}}, \tag{14.27}$$

inherits the uncertainty of $\dot{Q}_{\text{avail}}$ from Section 14.4.4 and of the required cooling $\dot{Q}_{\text{req}}$ from the forecast of the IT load. Take an illustrative plant with $\dot{Q}_{\text{avail}} = 12.0 \pm 0.6$ MW and

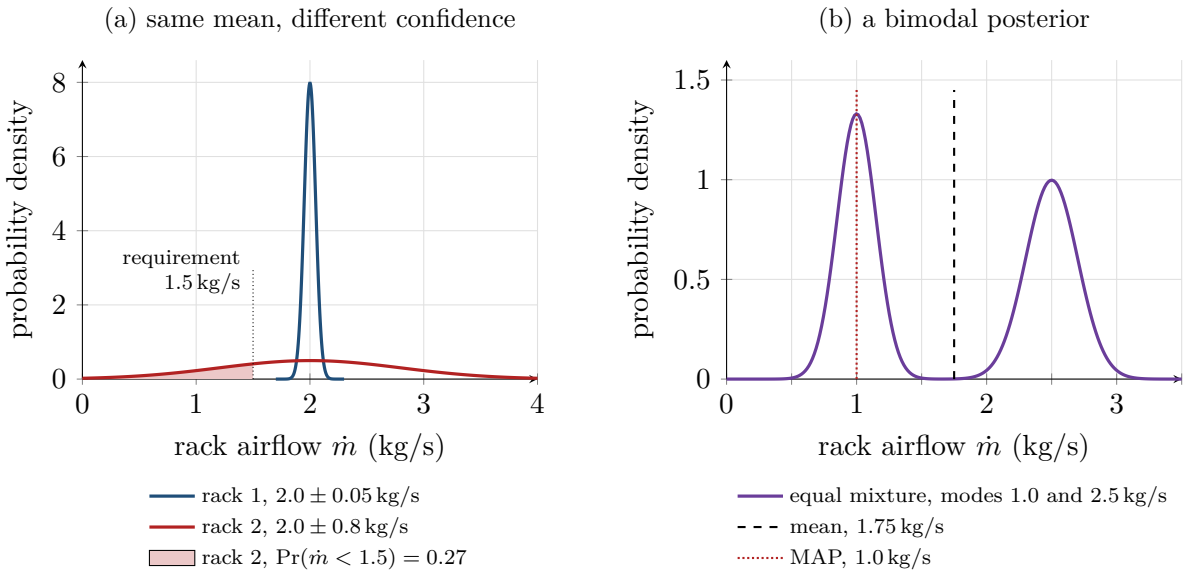


Figure 14.8: Two posteriors that a single number misrepresents (illustrative). (a) Normal posteriors with the same mean and standard deviations 0.05 and 0.8 kg/s; the shaded area is the probability that rack 2 is below a 1.5 kg/s requirement. (b) An equal mixture of $\mathcal{N}(1.0, 0.15^2)$ and $\mathcal{N}(2.5, 0.20^2)$: the mean lies where the density is near zero, and the MAP discards half the probability.

$\dot{Q}_{\text{req}} = 11.0 \pm 0.4$ MW for the peak hour, the two independent. The deterministic margin is $M = 1$ MW, which reads as safe. The margin's standard deviation is $\sqrt{0.6^2 + 0.4^2} = 0.72$ MW, and

$$\Pr(M < 0) = \Phi\left(\frac{0 - 1.0}{0.72}\right) = \Phi(-1.39) = 0.08.$$

The statement "the margin is 1 MW" becomes "there is an 8 % probability that the plant cannot meet the peak". For the single chiller of Section 14.4.4, 3.1 ± 0.3 MW against an illustrative known requirement of 2.1 MW, the margin is also 1 MW, but $\Pr(M < 0) = \Phi(-3.3) = 0.0004$. Equal margins carry risks that differ by a factor of two hundred.

Figure 14.9 draws the two margin distributions. The horizontal axis is the margin in megawatts and the vertical axis probability density. The solid curve is the plant, with the area below zero shaded; the dashed curve is the single chiller. Both are centered on 1 MW; the wide one puts visible probability below zero and the narrow one almost none. State estimation becomes risk estimation at this step, and the risk is a probability that available capacity is below required capacity. Chapter 17 develops capacity risk in full.

14.6.2 Estimation in the control loop

A controller chooses actions from a state. In its general form, an optimizing controller solves

$$u^* = \arg\min_u \; J(x, u), \tag{14.28}$$

where J is a cost such as energy plus a penalty on temperature violations. The state x in that problem is never known; the controller uses either the estimate $\hat{x}$ or the posterior $p(x \mid y)$. Using $\hat{x}$ treats the estimate as exact, which is adequate when the posterior is narrow. Using the posterior lets the controller minimize an expected cost or keep the probability of a violation below a tolerance, so that a rack with an uncertain airflow is cooled more conservatively than one with a well-determined airflow, as Section 14.5.4 requires.

The chain from data to action is therefore

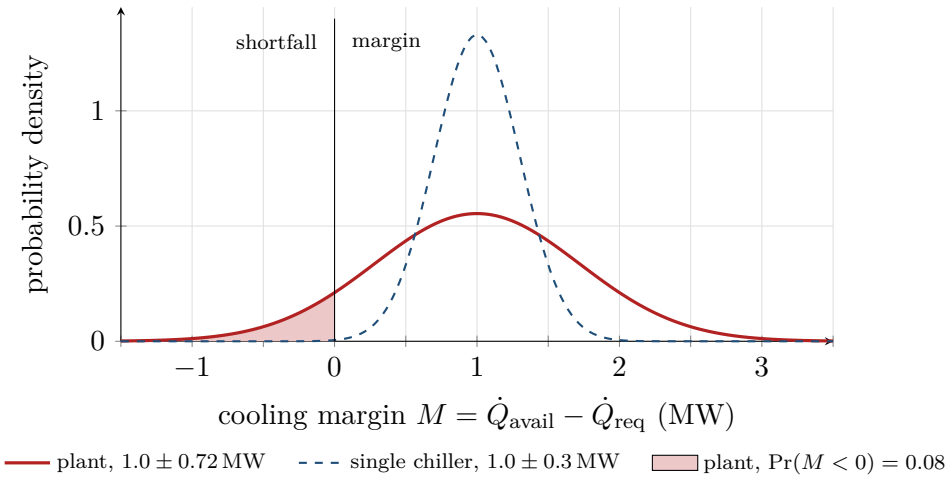


Figure 14.9: Distributions of the cooling margin for two cases with the same deterministic margin of 1 MW (illustrative): a plant with $\dot{Q}_{\text{avail}} = 12.0 \pm 0.6$ MW and $\dot{Q}_{\text{req}} = 11.0 \pm 0.4$ MW, and a single chiller with $\dot{Q}_{\text{avail}} = 3.1 \pm 0.3$ MW against a known 2.1 MW. The shaded area is the plant's probability of a shortfall.

$$\text{measurements } y \to \text{posterior } p(x \mid y) \to \text{risk } \Pr(\text{violation}) \to \text{action } u,$$

and each link depends on the one before it. A controller acting on a biased estimate acts consistently and wrongly; a controller acting on a correct estimate without its uncertainty cannot tell when to be cautious. The control loops of Chapter 10 and the optimization of Chapter 20 sit at the end of this chain.

14.6.3 Estimation, diagnosis and forecasting

Three neighboring tasks answer different questions about the same system, and Table 14.3 sets them side by side. State estimation answers what the current physical state is: the airflow of rack 7 is 1.8 kg/s. Diagnosis answers why the state is abnormal: the probability that the filter of rack 7 is blocked, given the readings, is 0.6. Forecasting answers what the state will be at a future time. Diagnosis takes the estimated state and its residuals as evidence and compares hypotheses H_i about causes; it needs the estimate but asks a different question (Chapter 13 and Chapter 16).

Forecasting propagates the current posterior forward through the process model,

$$p(x_{t+h} \mid y_{1:t}) = \int p(x_{t+h} \mid x_t)\, p(x_t \mid y_{1:t})\, \mathrm{d}x_t, \tag{14.29}$$

where $y_{1:t}$ denotes all readings up to time t and h is the horizon in samples. State estimation supplies the initial condition, and a better initial condition gives a better forecast at short horizons. For the filter of Figure 14.4, with $x_{k+1} - u = 0.9\,(x_k - u) + w_k$, the forecast standard deviation one minute ahead is 0.18 K when started from the filter's estimate ($\sigma = 0.16$ K) and 0.37 K when started from a raw reading ($\sigma = 0.4$ K); ten minutes ahead the two are 0.22 and 0.26 K, and thirty minutes ahead both approach the process's own spread of 0.23 K. The estimate matters most where the forecast is used most, in the next minutes of operation.

14.6.4 Filtering and smoothing

A filter estimates the state at time t from the readings up to t, $p(x_t \mid y_{1:t})$; it is what real-time operation needs, because the future readings do not yet exist. A smoother estimates the state at time t from a record that extends beyond it, $p(x_t \mid y_{1:T})$ with $T > t$. The later readings carry

Table 14.3: Four tasks that use the same model and measurements, and the question each answers.

Task	Object	Question	Example output
filtering (state estimation)	$p(x_t \mid y_{1:t})$	what is the physical state now?	rack 7 airflow 1.8 ± 0.1 kg/s
smoothing	$p(x_t \mid y_{1:T})$, $T > t$	what was the state then, using later data too?	airflow at the start of an event, for root-cause analysis
diagnosis	$\Pr(H_i \mid y)$	why is the state abnormal?	$\Pr(\text{blocked filter} \mid y) = 0.6$
forecasting	$p(x_{t+h} \mid y_{1:t})$	what will the state be?	inlet temperature in 10 minutes, 26.0 ± 0.22°C

information about the earlier state, because the state evolves continuously: a temperature that is high at minute 41 says something about minute 40 as well. A smoother runs a filter forward and then a correction backward through the record, and its estimates are at least as accurate as the filter's, usually more.

On the record of Figure 14.4 the effect is moderate: over minutes 20 to 119 the mean posterior standard deviation falls from 0.162 K for the filter to 0.140 K for the smoother, and the root-mean-square error from 0.133 to 0.124 K. The gain is larger for states that the readings constrain only indirectly, such as an airflow inferred through a slow thermal response, where a reading minutes later says more about the present than a reading now. Smoothing is the tool for post-event analysis: reconstructing what happened during an incident, root-cause analysis, and calibration of model parameters on historical data, where the whole record is available and accuracy matters more than immediacy.

14.6.5 Model adequacy

Every estimate is conditional on the model. Perfect sensors cannot save a wrong model, because the estimator's job is to find the state that makes the model agree with the data, and a wrong model agrees with the data only at a wrong state. Take a rack with true recirculation fraction $\alpha = 0.25$, $T_{\text{supply}} = 20$°C and $T_{\text{hot}} = 40$°C, so that its true inlet is 25°C, with $P_{\text{rack}} = 30$ kW and a true airflow of 2.0 kg/s, which makes its exhaust $25 + 15 = 40$°C. An estimator whose model omits recirculation assumes that the inlet is at the supply temperature, 20°C, and attributes the whole 20 K rise from supply to exhaust to the rack. It infers

$$\hat{\dot{m}} = \frac{30}{1.0 \times (40 - 20)} = 1.5\,\text{kg/s},$$

25 % low, with residuals of zero: the wrong model fits the readings exactly. If the airflow were fixed by a separate measurement, the same estimator would push the thermal resistance instead to an unrealistic value to reproduce the readings. The omitted path does not disappear; it is absorbed by whichever state or parameter is free to move.

The question before trusting an estimate is therefore whether the model is structurally adequate: whether it contains the paths, components and couplings that produce the behavior in the data. Residuals test this only partly. They expose a model error when the model cannot fit the data, but not when the error is absorbed by a free parameter, as in the example. Two checks help. A plausibility check compares the estimates with physical expectations: an airflow 25 % below the fan curve at the measured fan speed, or a thermal resistance outside the range of the equipment, signals that something is being absorbed. An independent measurement of the absorbing quantity, here a pressure-based airflow as in Section 14.1.5, exposes the conflict directly.

A good fit is not a correct explanation. A model with enough adjustable parameters reproduces a temperature record $T(t)$ well with parameters that are physically nonsensical: a negative capacitance in one node compensated by an excessive one in another, an airflow that exceeds the fan's maximum, a recirculation fraction above one. The fit is good because the parameters are many, not because the model is right. Fit quality is evidence of adequacy only when the parameters are few, identifiable and plausible. Estimation is therefore paired with three checks: plausibility of the estimated values, identifiability of the parameters from the data at hand, and structural adequacy of the model. Without them, a small residual is a statement about the fitting procedure, not about the data center.

14.6.6 A workflow for state estimation

The chapter's steps form a procedure, listed in Table 14.4. It begins by defining the physical states that the question needs and the measurement model that connects them to the readings; it checks observability before estimating, so that unobservable components are reported as such rather than given values; it estimates the state and quantifies its uncertainty; it checks the residuals and the plausibility of the estimates; and only then passes the posterior to diagnosis, risk and control. The order matters. An estimate computed before the observability check may be a number set by the prior; one passed downstream before the residual and plausibility check may be the output of a wrong model.

Table 14.4: A seven-step workflow for state estimation in a data center.

Step	Action	Content	Defined in
1	define physical states	x, θ for the question at hand	Section 14.1.6
2	define the measurement model	$y = h(x, \theta) + \epsilon$, noise covariances	Sections 14.1.2, 14.2.1
3	check observability	which components the readings determine	Section 14.5.1
4	estimate the state	filter, smoother or weighted least squares	Sections 14.2.4, 14.3.1
5	quantify uncertainty	covariance or full posterior	Section 14.5.4
6	check residuals and plausibility	bad data, model adequacy	Sections 14.3.4, 14.6.5
7	propagate	to diagnosis, risk and control	Sections 14.6.1, 14.6.2

14.6.7 The estimation equations

Five equations carry the chapter:

$$\text{process model:} \quad x_{k+1} = f(x_k, u_k, d_k, \theta) + w_k, \tag{14.30}$$

$$\text{measurement model:} \quad y_k = h(x_k) + v_k, \tag{14.31}$$

$$\text{filtered posterior:} \quad p(x_k \mid y_{1:k}), \tag{14.32}$$

$$\text{residual:} \quad r_k = y_k - h(\hat{x}_k), \tag{14.33}$$

$$\text{MAP estimate:} \quad x_{\text{MAP}} = \arg\max_x \; p(x \mid y). \tag{14.34}$$

The first two are the physics and the sensors; the third is what an estimator computes; the fourth is how it is tested; the fifth is the single number it reports when one is required, with the loss of information that Section 14.5.5 describes.

The broader statement is that a data center does not know x; it sees y. Its operation rests on a belief about the physical state, formed from measurements through a model. The operational question is therefore not "what does the sensor say" but "given all the sensors, the physics, the

uncertainty and the topology, what physical state is plausible, and how plausible". Diagnosis, risk and control are downstream of that belief, and they are only as sound as the probabilistic physical model that produces it.

14.7 What to retain

Definition 14.1 (State estimation)**.** State estimation infers the hidden physical state of a data center from incomplete, noisy measurements by combining a process model, a measurement model and the readings into a posterior $p(x \mid y)$, whose center is the estimate and whose width and shape are its uncertainty. It is valid only where the measurements make the state observable and the model is structurally adequate.

Table 14.5 lists the chapter's objects and the subsection that defines each.

Table 14.5: The objects of state estimation and where each is defined.

Object	Expression	Role	Defined in
Posterior	$p(x \mid y) \propto p(y \mid x)\, p(x)$	the complete answer	14.1.2
Rack airflow	equation (14.3)	an inferred state	14.1.3
Propagated uncertainty	2.0 ± 0.12 to $\pm 0.15\,\mathrm{kg/s}$	the width of an estimate	14.1.4
Redundant evidence	equation (14.4)	agreement or conflict	14.1.5
Augmented state	$x_{\mathrm{aug}} = [x;\ \theta]$	tracks slow parameters	14.1.7
Process model	$x_{k+1} = f(x_k, u_k, d_k, \theta) + w_k$	what the physics predicts	14.2.1
Innovation	$r_k = y_k - h(\hat{x}_{k\mid k-1})$	the unexpected part of a reading	14.2.2
Kalman gain	$K_k = \sigma^2_{k\mid k-1}/(\sigma^2_{k\mid k-1} + \sigma^2_v)$	model against sensor	14.2.4
Weighted least squares	equation (14.18)	a snapshot estimate	14.3.1
Residual test	$\chi^2 = \sum_j r_j^2/\sigma_j^2$	bad data and model errors	14.3.3
Recirculation fraction	equation (14.21)	an inferred parameter	14.4.1
Coil conductance	$\widehat{UA} = \dot{Q}/\Delta T_{\mathrm{lm}}$	condition estimation	14.4.2
Available capacity	$\dot{Q}_{\mathrm{avail}} = 3.1 \pm 0.3\,\mathrm{MW}$	a state, not a rating	14.4.4
Uncertain topology	$p(x, G \mid y)$	structural inference	14.4.5
Observability	$h(x_1) = h(x_2),\ x_1 \neq x_2$	what the data can determine	14.5.1
Excitation	$\tau = C_{\mathrm{th}}/(\dot{m} c_{p,\mathrm{air}})$	transients add information	14.5.3
MAP estimate	$\arg\max_x p(x \mid y)$	one number, shape lost	14.5.5
Margin risk	$\Pr(\dot{Q}_{\mathrm{avail}} - \dot{Q}_{\mathrm{req}} < 0)$	estimation becomes risk	14.6.1
Smoothing	$p(x_t \mid y_{1:T}),\ T > t$	post-event analysis	14.6.4

14.8 Exercises

Exercise 14.1. A rack draws $P_{\mathrm{rack}} = 24 \pm 0.8\,\mathrm{kW}$, with $T_{\mathrm{in}} = 21°\mathrm{C}$ and $T_{\mathrm{out}} = 33°\mathrm{C}$, each temperature $\pm 0.4\,\mathrm{K}$; take $c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ/(kg\,K)}$. (a) Estimate the rack airflow with equation (14.3) and its standard deviation. (b) A pressure route gives $1.85 \pm 0.12\,\mathrm{kg/s}$. Compute the difference between the two estimates in standard deviations and, if they agree, their inverse-variance combination. (c) Repeat (b) for a pressure estimate of $2.6 \pm 0.12\,\mathrm{kg/s}$, and list two explanations for the result.

Exercise 14.2. A scalar Kalman filter tracks a hall temperature. The previous estimate is $25.0°\mathrm{C}$ with variance $0.09\,\mathrm{K}^2$; the model predicts $25.2°\mathrm{C}$ and adds a process variance of $0.03\,\mathrm{K}^2$; the sensor variance is $0.25\,\mathrm{K}^2$. (a) Compute the predicted variance, the gain and the corrected estimate and variance for a reading of $25.8°\mathrm{C}$. (b) Compute the normalized innovation for a

reading of 29.0°C; should the filter use it under a four-sigma gate? (c) The sensor is replaced by one with variance $0.01\,\mathrm{K}^2$. Recompute the gain for the reading of (a) and state which source the estimate now follows.

Exercise 14.3. Four inlet sensors with standard deviation 0.25 K read 22.8, 23.1, 22.9 and 27.4°C, and the model predicts 23.0°C for all four. (a) Compute the normalized residuals against the model and identify the bad reading. (b) Compute the common estimate from all four readings and the normalized residuals against it; which sensors now exceed three standard deviations? (c) Recompute the estimate without the bad reading. (d) State three physical explanations for the bad reading and one measurement that would separate two of them.

Exercise 14.4. (a) A rack has $T_{\mathrm{supply}} = 18°\mathrm{C}$, $T_{\mathrm{hot}} = 38°\mathrm{C}$ and $T_{\mathrm{in}} = 23°\mathrm{C}$, each $\pm 0.5\,\mathrm{K}$. Compute α and its standard deviation. (b) A coil transfers $\dot{Q} = 250\,\mathrm{kW}$ with air entering at 30°C and leaving at 19°C and water entering at 13°C and leaving at 19°C, in counterflow. Compute ΔT_{lm} and $\widehat{UA}$. (c) A plant has $\dot{Q}_{\mathrm{avail}} = 8.0 \pm 0.5\,\mathrm{MW}$ and $\dot{Q}_{\mathrm{req}} = 7.2 \pm 0.3\,\mathrm{MW}$, independent. Compute the deterministic margin, its standard deviation and $\Pr(M < 0)$. (d) A second plant has the same deterministic margin with $\sigma_M = 0.2\,\mathrm{MW}$. Compute its shortfall probability and say which plant needs action first.

14.9 Connection to the TERRA framework

State estimation is the inference half of a graph model: it turns the conservation networks of the earlier chapters and the measurements attached to them into a posterior over everything that is not measured.

- **The state vector lives on nodes and edges.** The temperatures of equation (14.5) are states of aisle, rack and plenum nodes in the thermal network; the rack airflows and the water flow are flows on edges of the air and water networks; the plenum pressure is a node potential. The process model of equation (14.9) is forward simulation of those networks in time, and the measurement model of equation (14.10) attaches each reading to a node or an edge with its noise.

- **Parameters are component properties, inferred with the state.** The coil conductance, flow coefficients, thermal resistances and recirculation fraction of equation (14.6) are properties of coil, restriction and mixing components. Appending them to the state, as in equation (14.8), is the same posterior inference applied to slowly varying quantities, which is how degradation such as the falling UA of Table 14.2 is tracked.

- **Balances that should close are residual generators.** The two airflow estimates of Figure 14.1 and the three readings of Table 14.1 are redundancy relations of the air network; their residuals, tested as in Section 14.3.3, flag bad data and model errors. Rack airflow, recirculation fraction and coil conductance come out as inferred quantities with posteriors, kept apart from what is measured and what is calculated.

- **Topology is an uncertain object.** Which cooling unit serves which row is an edge set of the graph, and equation (14.23) weighs candidate edge sets by how well each explains the data, as in Figure 14.7.

- **Observability is checked on the graph, and the posterior feeds risk.** Whether the attached measurements determine a node or edge quantity is a property of the graph and the sensor attachments, checked before estimation; and the posterior over available and required cooling gives the risk of Figure 14.9, a probability that available capacity is below required capacity.

In summary, a data center's measurements attach to a few nodes and edges of its coupled electrical, air-and-water and thermal networks; state estimation combines them with the networks' conservation laws and component models into a posterior over the hidden states, parameters and topology; residuals test that posterior against the data, observability decides in advance which parts of it the data can narrow, and risk, diagnosis and control are computed from it. The next chapter, Chapter 15, examines that posterior's width directly: how uncertainty is quantified and when the data determine the parameters at all.

Chapter 15

Uncertainty and Identifiability

15.1 Kinds of uncertainty and the meaning of identifiability

Chapter 14 estimated the unmeasured states and parameters of a data center from its measurements: a rack airflow from a power and a temperature difference, a thermal resistance from a transient, a sensor bias from a balance that does not close. Those estimates came with uncertainty, and the uncertainty shrank as more and better data arrived. This chapter examines a different possibility: that the data do not merely leave a quantity uncertain but fail, by their nature, to distinguish between several explanations of what was observed. The question of the chapter is

when is a quantity merely uncertain, and when does the available data fundamentally fail to distinguish between several explanations of the same evidence?

The distinction is practical. It decides which of four things an analysis should do with a question: infer an answer, hedge it with a wide uncertainty, ask for more data of a named kind, or state that the data cannot answer it. The chapter first separates the kinds of uncertainty and defines identifiability; it then describes how identifiability appears in the geometry of a posterior distribution, how it limits diagnosis, how priors and model structure can disguise it, and how it is judged against the decision the inference is meant to support.

Jargon: identifiability, structural and practical identifiability, likelihood, prior, posterior, excitation, Fisher information, flat direction, model evidence, observational equivalence

A quantity is *identifiable* from a set of measurements when different values of it produce different predictions of those measurements, so that in principle the measurements single out one value. *Structural* identifiability asks this of perfect, noise-free data from the given sensors and inputs; *practical* identifiability asks it of the actual, finite and noisy record. The *likelihood* $p(y \mid \theta)$ is the probability density of the measurements y for given parameter values θ, read as a function of θ. The *prior* $p(\theta)$ is the distribution of the parameters before the measurements are used, and the *posterior* $p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta)$ is their distribution after. *Excitation* is variation of the inputs of a system, such as a change of rack power, that makes its dynamics visible in the measurements. The *Fisher information* $\mathcal{I}(\theta)$ is a matrix that measures how sharply the likelihood changes as the parameters move; its eigenvectors with small eigenvalues are *flat directions*, combinations of parameters along which the data barely change. The *model evidence* $p(y \mid \mathcal{M}_j)$ is the probability of the measurements under a candidate model structure $\mathcal{M}_j$, averaged over that model's parameters. Two model structures are *observationally equivalent* for a sensor set when they predict the same distribution of every measurement that set can take.

Key idea

Uncertainty is a width; non-identifiability is a shape. More and better data of the same kind shrink the width of an estimate. They do nothing to a set of explanations that fit the evidence equally well, which yields only to independent information: a different sensor, a varied input, a physical constraint. When none is available, the correct result of an analysis is the statement that the data cannot decide, together with the measurement that would.

15.1.1 Four kinds of uncertainty

Uncertainty in a data center model enters at four levels, and they differ in what causes them and in what reduces them.

Measurement uncertainty. A sensor reports the quantity it measures with an error,

$$y = x + \epsilon, \tag{15.1}$$

where x is the true value and ϵ the random error. Chapter 9 described its sources. A rack inlet sensor that reads $T = 24.0 \pm 0.5°\mathrm{C}$ leaves the true inlet temperature anywhere in a band about a degree wide.

Parameter uncertainty. The constants of the model are not known exactly. A rack's thermal resistance R_{th} and a cooling coil's overall conductance UA are described by distributions, $R_{\mathrm{th}} \sim p(R_{\mathrm{th}})$ and $UA \sim p(UA)$, which the data can narrow.

State uncertainty. Some quantities that change with time are not observed directly. The airflow through a rack, $\dot{m}_{\mathrm{rack}}$, and the recirculation fraction α of its inlet air are states of the airflow network that no sensor reads in most facilities; they are inferred from the temperatures and powers that are read.

Structural uncertainty. The model structure itself may be wrong: an airflow path the model does not contain, an undocumented piping connection, a heat source that is missing from the balance, an incorrect mapping of racks to power distribution units. This kind is often the most dangerous of the four. The other three are quantified *within* a model, and an analysis reports them as error bars; a wrong structure is not quantified by the model it invalidates, and it produces estimates that are precise, internally consistent and wrong.

Figure 15.1 arranges the four kinds. The dashed outer box is the model structure $\mathcal{M}$, the topology G and the equations written on it; uncertainty about the box itself is structural. Inside it, from left to right, are the parameters θ, the states x they govern, and the sensor relation that turns states into readings. The box on the right, outside the model, is the record of measurements y. Each inner box is labeled with its kind of uncertainty and an example. The arrows show the direction of prediction: parameters determine states, states determine readings. Inference runs the other way, from the record back into the boxes, and it can only reach what the arrows carry.

15.1.2 Uncertainty that more data reduces

Most of the uncertainty in the first three kinds can be reduced. A better sensor narrows the measurement error, for example from $\pm 0.5°\mathrm{C}$ to $\pm 0.1°\mathrm{C}$. More data narrow the distribution of a parameter: if each of N independent readings carries a random error with standard deviation σ, their mean carries $\sigma/\sqrt{N}$, so a hundred readings of a sensor with $\sigma = 0.5\,\mathrm{K}$ fix a constant

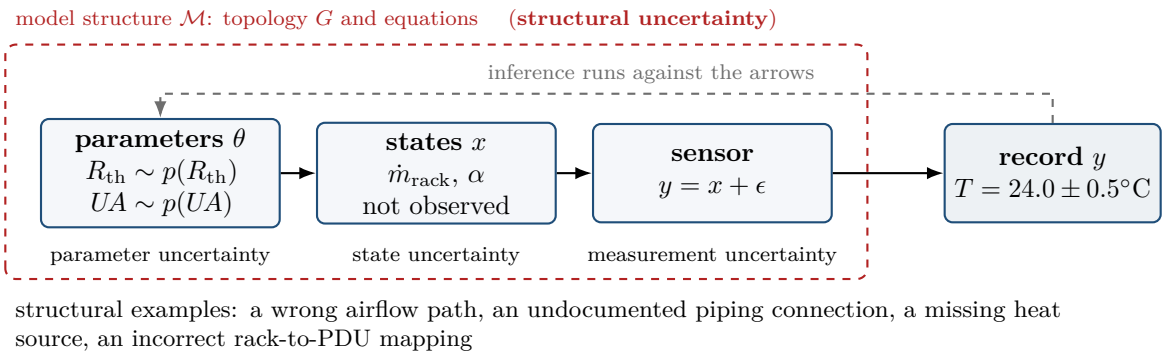


Figure 15.1: Four kinds of uncertainty. Parameters, states and sensor errors are uncertain within a model and can be given distributions; the model structure itself (dashed box) can be wrong, and that error is not reported by the model's own error bars. Examples are illustrative.

temperature to 0.05 K. In the language of inference the posterior $p(R_{\text{th}} \mid y)$ becomes narrower as the record grows, and with enough data it concentrates around one value.

This is the ordinary behavior of an estimate, and it is the behavior the methods of Chapter 14 rely on. The rest of the chapter concerns a different behavior, in which the posterior does *not* concentrate around one value however much data of the given kind arrives. That behavior is a problem of identifiability, and it is not cured by the remedies that cure uncertainty.

15.1.3 Identifiability as uniqueness of explanation

The air-side energy balance of Chapter 5 relates the heat carried by a stream of air to its mass flow and temperature rise,

$$\dot{Q} = \dot{m}\, c_p\, \Delta T, \tag{15.2}$$

with $c_p \approx 1.0\,\text{kJ/(kg K)}$ for air. Consider two situations at one rack.

In the first, the heat is known and the temperature rise is measured: $\dot{Q} = 20\,\text{kW}$ and $\Delta T = 10\,\text{K}$. The heat is not itself measured; it is calculated from the rack's measured electrical power, which leaves the rack as heat. The airflow then follows uniquely,

$$\dot{m} = \frac{\dot{Q}}{c_p\, \Delta T} = \frac{20\,\text{kW}}{1.0\,\text{kJ/(kg K)} \times 10\,\text{K}} = 2\,\text{kg/s},$$

and it is identifiable: one value of $\dot{m}$, and only one, reproduces the data.

In the second, only the temperature rise is measured, $\Delta T = 10\,\text{K}$. Every pair with $\dot{Q}/\dot{m} = c_p\, \Delta T = 10\,\text{kJ/kg}$ fits:

$$(\dot{Q}, \dot{m}) = (10\,\text{kW}, 1\,\text{kg/s}),\ (20\,\text{kW}, 2\,\text{kg/s}),\ (40\,\text{kW}, 4\,\text{kg/s}),\ \ldots$$

The heat and the airflow are not separately identifiable. Their ratio is identifiable; each factor is not. Formally, a parameter vector θ is identifiable from measurements y when

$$p(y \mid \theta_1) = p(y \mid \theta_2) \text{ for every } y \quad \Longrightarrow \quad \theta_1 = \theta_2, \tag{15.3}$$

that is, when no two different parameter values predict the same distribution of measurements. In the second situation $\theta_1 = (10\,\text{kW}, 1\,\text{kg/s})$ and $\theta_2 = (20\,\text{kW}, 2\,\text{kg/s})$ predict the same ΔT for every noise realization, and the implication fails.

Figure 15.2 draws both situations in the plane of the two unknowns. The horizontal axis is the airflow $\dot{m}$ in kg/s; the vertical axis is the heat $\dot{Q}$ in kW. The thick line is the set of pairs consistent with an exact reading $\Delta T = 10\,\text{K}$; the shaded wedge around it is the set consistent

with a reading of $10 \pm 0.5\,\mathrm{K}$. The three dots are the pairs listed above. The horizontal green band is the independent constraint of the first situation, a rack power of $20 \pm 1\,\mathrm{kW}$ taken as the heat. The temperature reading alone confines the answer to a line or a wedge that runs to infinity; only where the green band crosses it is the answer bounded.

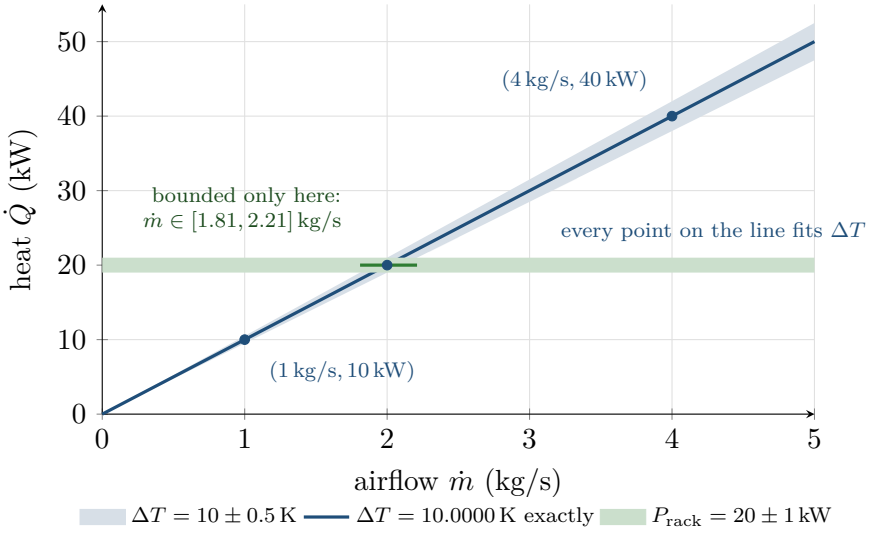


Figure 15.2: Pairs of heat and airflow consistent with a rack temperature rise, $\dot{Q} = \dot{m}\, c_p\, \Delta T$ with $c_p = 1.0\,\mathrm{kJ/(kg\,K)}$ (equation (15.2)). An exact reading (line) and a noisy reading (wedge) both leave the pair unbounded; an independent power measurement (green band) bounds it. Values illustrative.

15.1.4 Precision and independent information

Better precision does not solve the problem. Suppose the temperature rise were measured as $\Delta T = 10.0000\,\mathrm{K}$ with almost no noise. The wedge of Figure 15.2 collapses onto its center line, but the line still runs from the origin to infinity, and the pairs $(10\,\mathrm{kW}, 1\,\mathrm{kg/s})$ and $(20\,\mathrm{kW}, 2\,\mathrm{kg/s})$ remain exactly as consistent with the reading as before. A more precise sensor of the same quantity narrows the set of explanations in the direction across the line and leaves it unbounded along the line. Better precision is not better identifiability.

What bounds the set is an additional, independent constraint, one that cuts across the line instead of running along it. The rack power does this: since nearly all the electrical power of a rack leaves as heat, $P_{\mathrm{rack}} \approx \dot{Q}$, and then

$$\dot{m} = \frac{P_{\mathrm{rack}}}{c_p\, \Delta T}. \tag{15.4}$$

With $P_{\mathrm{rack}} = 20 \pm 1\,\mathrm{kW}$ and $\Delta T = 10 \pm 0.5\,\mathrm{K}$ the airflow lies between $19/(1.0 \times 10.5) = 1.81\,\mathrm{kg/s}$ and $21/(1.0 \times 9.5) = 2.21\,\mathrm{kg/s}$: uncertain, but bounded and centered on $2\,\mathrm{kg/s}$. "Independent" has a precise meaning here. A second temperature sensor on the same air stream is not independent in this sense: it measures the same ΔT and constrains the same ratio. A measurement adds identifiability when it constrains a combination of the unknowns that the existing measurements do not.

15.1.5 Samples and independent constraints

More samples are not more information of this kind either. A record of 10 000 readings of $\Delta T = 10\,\mathrm{K}$, taken while nothing in the rack changes, repeats one constraint ten thousand times:

$\dot{Q}/\dot{m} = 10\,\mathrm{kJ/kg}$. The average of the readings is very precise, about $0.5/\sqrt{10\,000} = 0.005\,\mathrm{K}$ for a sensor with a 0.5 K random error, but it is a precise value of the same ratio. In linear-algebra terms, each reading contributes one row to the system of equations that relates measurements to unknowns, and ten thousand identical rows have rank one. Two unknowns need rank two.

The distinction between samples and constraints recurs throughout the chapter. A long record of steady operation is many samples of a few constraints. A short record that contains a change of load, a step of fan speed or a failed unit can hold more independent constraints than a year of steady data, because the change makes the unknowns enter the measurements in a new combination.

Principle 15.1. *Precision and sample count reduce uncertainty within an identifiable model; neither creates identifiability. Only an independent constraint, a different sensor, a varied input or a physical relation that cuts across the existing ones, separates explanations that fit the data equally well.*

15.2 Parameter combinations and posterior geometry

Non-identifiability rarely means that a parameter is entirely unknown. More often the data fix a combination of parameters and leave the individual parameters free along a curve. This section shows how input variation separates parameters, describes the resistance–capacitance ambiguity and the shape it gives a posterior, and introduces the information matrix, whose eigenvalues and eigenvectors make the shape quantitative.

15.2.1 Dynamic excitation

The first-order thermal model of Chapter 8 describes a rack or zone with thermal resistance R_{th} and thermal capacitance C_{th}, heated by the rack power and cooled toward a coolant temperature T_{cool}:

$$C_{\mathrm{th}}\,\frac{\mathrm{d}T}{\mathrm{d}t} = P_{\mathrm{rack}} - \frac{T - T_{\mathrm{cool}}}{R_{\mathrm{th}}}, \qquad \tau = R_{\mathrm{th}}C_{\mathrm{th}}. \tag{15.5}$$

In steady operation $\mathrm{d}T/\mathrm{d}t = 0$ and the model reduces to $T = T_{\mathrm{cool}} + R_{\mathrm{th}}P_{\mathrm{rack}}$. The capacitance has dropped out: a steady record determines R_{th}, given the power and the coolant temperature, and says nothing about C_{th}. When $P_{\mathrm{rack}}(t)$ varies, the storage term is no longer zero, the temperature moves from one equilibrium to the next with time constant τ, and the timing of that movement carries C_{th}. A sufficiently varied input lets the trajectory determine both parameters; steady operation hides the parameters that transients reveal.

Figure 15.3 shows this with three illustrative parameter sets. The upper panel is the rack power against time in minutes: 6 kW, then 9 kW from minute 30, then 5 kW from minute 90. The lower panel is the rack temperature, with $T_{\mathrm{cool}} = 18\,^{\circ}\mathrm{C}$. Set A has $R_{\mathrm{th}} = 1\,\mathrm{K/kW}$ and $C_{\mathrm{th}} = 600\,\mathrm{kJ/K}$, so $\tau = 600\,\mathrm{s} = 10\,\mathrm{min}$. Set A′ has the same resistance and twice the capacitance, $\tau = 20\,\mathrm{min}$. Set B has twice the resistance and half the capacitance, $R_{\mathrm{th}} = 2\,\mathrm{K/kW}$ and $C_{\mathrm{th}} = 300\,\mathrm{kJ/K}$, and therefore the same time constant as A. On the plateaus A and A′ coincide, at 24, 27 and 23 °C: only the transients after the two steps tell them apart, and they do so for about half an hour after each step. B sits on plateaus twice as far above the coolant, 30, 36 and 28 °C, and moves with exactly the timing of A. The plateaus separate A from B; the transients separate A from A′. A record with both separates all three.

15.2.2 The resistance–capacitance ambiguity

The time constant is a product. The pairs

$$(R_{\mathrm{th}}, C_{\mathrm{th}}) = (1\,\mathrm{K/kW},\ 600\,\mathrm{kJ/K}) \quad \text{and} \quad (2\,\mathrm{K/kW},\ 300\,\mathrm{kJ/K})$$

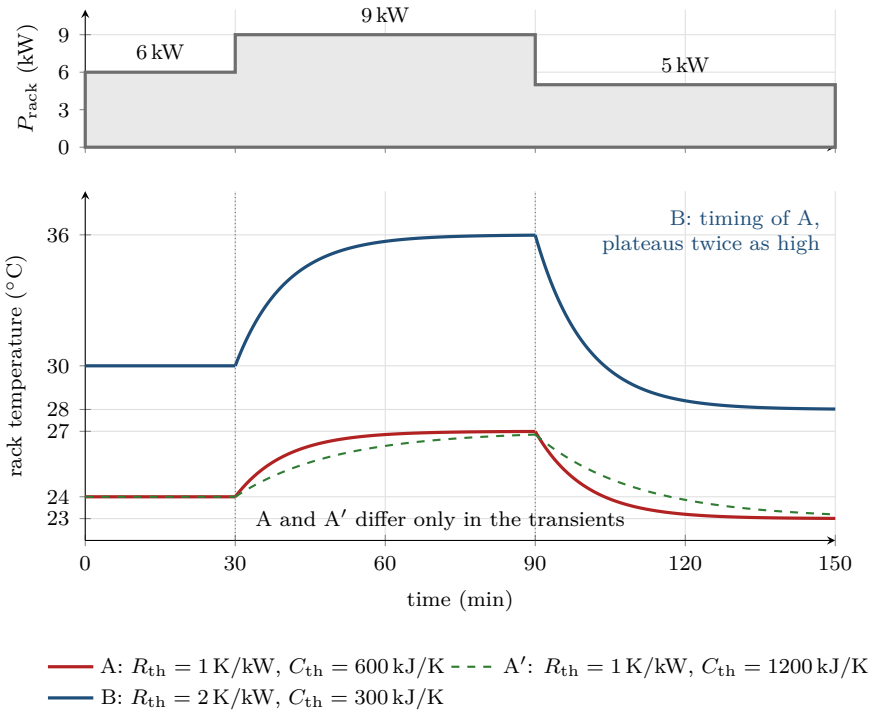


Figure 15.3: Excitation and identifiability for the first-order model $C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = P_{\text{rack}} - (T - T_{\text{cool}})/R_{\text{th}}$ (equation (15.5)), $T_{\text{cool}} = 18°\text{C}$, under an illustrative stepped rack power. Plateaus $T = T_{\text{cool}} + R_{\text{th}}P_{\text{rack}}$ fix R_{th}; the exponential approaches with time constant $\tau = R_{\text{th}}C_{\text{th}}$ (10 min for A and B, 20 min for A′) fix the capacitance once R_{th} is known.

give the same $\tau = R_{\text{th}}C_{\text{th}} = 600\,\text{s}$, since $1\,\text{K/kW} \times 600\,\text{kJ/K} = 600\,\text{s}$. If the equilibrium temperature also informs R_{th}, as the plateaus of Figure 15.3 do when the heat reaching the measured volume is known, the two pairs separate. If it does not, because only part of the rack power reaches the measured air, or the size of a step was not recorded, or the record shows only the normalized shape of a response, then the data determine τ and nothing else: Chapter 8 showed two responses with the same time constant that coincide once each is divided by its final change.

The general statement is that a model may be identifiable in some combinations of its parameters and not in the parameters themselves. When the predicted measurements depend on θ only through a function $\phi(\theta)$,

$$p(y \mid \theta) = g(y, \phi(\theta)), \tag{15.6}$$

then ϕ is the most that can be identified, and every θ on a level set of ϕ is an equally good explanation. In the example $\phi = R_{\text{th}}C_{\text{th}}$; in Section 15.1.3 it was $\dot{Q}/\dot{m}$. The *identifiable combination* is a property of the model and the measurements together, and recognizing it is often the most useful single result of an identifiability analysis: it says what to report.

15.2.3 Posterior geometry and correlation

The ambiguity has a shape. A posterior over $(R_{\text{th}}, C_{\text{th}})$ computed from data that fix τ is not a compact cloud around one point but a long ridge along the curve $R_{\text{th}}C_{\text{th}} \approx 600\,\text{s}$: an elongated valley of the negative log-posterior rather than a round bowl. In logarithmic coordinates the curve becomes a straight line, $\ln R_{\text{th}} + \ln C_{\text{th}} = \ln \tau$, and the ridge becomes an elongated ellipse oriented across both axes. The two parameters are strongly coupled; only their product is well constrained.

Figure 15.4 draws two such posteriors. Both axes are logarithmic: the horizontal axis is R_{th} in K/kW, the vertical axis is C_{th} in kJ/K. The thin gray ridge is the region holding about 95 % of the posterior when the data fix $\ln \tau$ to a standard deviation of 0.02 (the time constant to 2 %) and the individual parameters are held only by a broad prior, a standard deviation of 0.7 on each logarithm, about a factor of two. It runs along the dotted line $R_{\mathrm{th}}C_{\mathrm{th}} = 600\,\mathrm{s}$ from below 0.4 to above 2.6 K/kW and contains both pairs of Section 15.2.2, the two marked points. The small red ellipse is the same posterior after an equilibrium temperature adds information on R_{th} alone, a standard deviation of 0.1 on $\ln R_{\mathrm{th}}$: it is centered on pair A and excludes pair B.

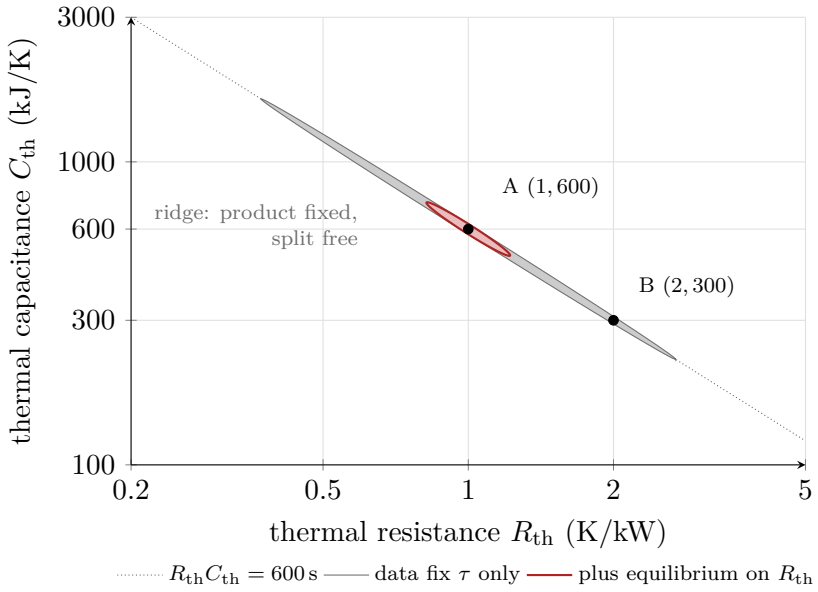


Figure 15.4: Posterior regions (about 95 %) for thermal resistance and capacitance, Gaussian in logarithmic coordinates, illustrative. Gray: data fix $\ln(R_{\mathrm{th}}C_{\mathrm{th}})$ to a standard deviation of 0.02, with a prior standard deviation of 0.7 on each logarithm; correlation -0.999. Red: an equilibrium adds a standard deviation of 0.1 on $\ln R_{\mathrm{th}}$; correlation -0.98, yet each parameter is known to about 10 %.

Correlation is a warning, not a proof. The correlation coefficient of two parameters in the posterior,

$$\rho(R_{\mathrm{th}}, C_{\mathrm{th}}) = \frac{\mathrm{Cov}(R_{\mathrm{th}}, C_{\mathrm{th}})}{\sqrt{\mathrm{Var}(R_{\mathrm{th}})\ \mathrm{Var}(C_{\mathrm{th}})}}, \tag{15.7}$$

measures how strongly they trade against each other. A value such as $\rho = -0.98$ indicates a strong trade-off, but it is not a proof of non-identifiability. The two regions of Figure 15.4 make the point: the gray ridge has a correlation of -0.999 in the logarithms and leaves each parameter uncertain by a factor of about two; the red ellipse has -0.98 and fixes each parameter to about 10 %, which is narrow enough for most purposes. Correlation describes the orientation of the posterior, not its size. A judgement needs the posterior widths, the covariance matrix, its eigenvalues and eigenvectors (the next subsection), the profile likelihood, and the sensitivity of the predictions to each parameter. The *profile likelihood* of one parameter is the likelihood maximized over all the others at each fixed value of that parameter; a profile that stays flat over a wide range shows that the data cannot fix the parameter even when the others are free to adjust in its favor.

15.2.4 Flat directions and the information matrix

The quantitative tool is the Fisher information matrix. For a parameter vector θ and a likelihood $p(y \mid \theta)$ it is

$$\mathcal{I}(\theta) = \mathbb{E}\Big[\nabla_\theta \ln p(y \mid \theta)\; \nabla_\theta \ln p(y \mid \theta)^{\mathsf{T}}\Big], \tag{15.8}$$

the expected curvature of the log-likelihood. For measurements with Gaussian errors of covariance Σ_ϵ and a sensitivity matrix $J = \partial\hat{y}/\partial\theta$ of the predicted measurements $\hat{y}$, it is $\mathcal{I} = J^{\mathsf{T}}\Sigma_\epsilon^{-1}J$. Near the best estimate the posterior is approximately Gaussian with covariance $(\mathcal{I} + \Sigma_0^{-1})^{-1}$, where Σ_0 is the prior covariance: information from the data and from the prior add.

The eigendecomposition $\mathcal{I} = \sum_i \lambda_i v_i v_i^{\mathsf{T}}$ reads the matrix by direction. Along an eigenvector v_i the log-likelihood falls off with curvature λ_i, so that, with a flat prior, the posterior standard deviation along v_i is about $1/\sqrt{\lambda_i}$. Large eigenvalues are strongly constrained directions; small ones are weakly constrained; an eigenvalue of zero is a direction along which the likelihood does not change at all, and a nearly singular $\mathcal{I}$ signals poorly identifiable combinations. The ratio $\lambda_{\max}/\lambda_{\min}$, the condition number, says how elongated the posterior is.

A worked 2×2 example. Work in the logarithms $\tilde{\theta} = (\ln R_{\mathrm{th}}, \ln C_{\mathrm{th}})$, which measure relative changes. A transient that fixes $\ln\tau = \ln R_{\mathrm{th}} + \ln C_{\mathrm{th}}$ with a standard deviation of 0.02 has sensitivity row $J = (1, 1)$ and noise variance 0.02^2, so

$$\mathcal{I}_\tau = \frac{1}{0.02^2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2500 & 2500 \\ 2500 & 2500 \end{pmatrix}.$$

Its eigenvalues are $\lambda_1 = 5000$ and $\lambda_2 = 0$, with eigenvectors $v_1 = (1, 1)/\sqrt{2}$, the product direction, and $v_2 = (1, -1)/\sqrt{2}$, the split direction. The zero eigenvalue is the ridge of Figure 15.4: the data say nothing about $\ln R_{\mathrm{th}} - \ln C_{\mathrm{th}}$. Adding an equilibrium that fixes $\ln R_{\mathrm{th}}$ with a standard deviation of 0.1 adds $1/0.1^2 = 100$ to the upper-left entry:

$$\mathcal{I} = \begin{pmatrix} 2600 & 2500 \\ 2500 & 2500 \end{pmatrix}, \qquad \lambda_{1,2} = \frac{5100 \pm \sqrt{5100^2 - 4\times 250\,000}}{2} = 5050.5,\; 49.5,$$

with trace 5100 and determinant $2600 \times 2500 - 2500^2 = 250\,000$. The weak eigenvector is now $v_2 \approx (0.70, -0.71)$. The inverse,

$$\mathcal{I}^{-1} = \frac{1}{250\,000}\begin{pmatrix} 2500 & -2500 \\ -2500 & 2600 \end{pmatrix} = \begin{pmatrix} 0.0100 & -0.0100 \\ -0.0100 & 0.0104 \end{pmatrix},$$

gives standard deviations of 0.10 on $\ln R_{\mathrm{th}}$ and 0.102 on $\ln C_{\mathrm{th}}$, about 10 % on each parameter, a correlation of $-0.0100/(0.10 \times 0.102) = -0.98$, and a standard deviation of $1/\sqrt{49.5} = 0.14$ along the split direction. These are the numbers behind the red ellipse. The condition number fell from infinite to $5050.5/49.5 \approx 102$: still elongated, no longer degenerate.

Three parameters. With $\theta = (R_{\mathrm{th}}, C_{\mathrm{th}}, \alpha)$, resistance, capacitance and recirculation fraction, all scaled as in Section 15.2.5, a typical result is $\lambda_1 \gg \lambda_2 \gg \lambda_3$, with a very small λ_3 and an eigenvector such as

$$v_3 = 0.7\,\tilde{R_{\mathrm{th}}} - 0.7\,\tilde{C_{\mathrm{th}}} + 0.1\,\tilde{\alpha}.$$

It says that raising the resistance while lowering the capacitance by the same scaled amount, with the recirculation fraction nearly unchanged, barely changes the likelihood. That is more informative than the three individual standard deviations, which would all be large without saying that their errors are tied together in one specific combination, or which combination it is.

15.2.5 Scaling

Eigenvectors depend on the units of the parameters. In raw units a room-scale model might have $R_{\mathrm{th}} = 0.001\,\mathrm{K/W}$ and $C_{\mathrm{th}} = 10^8\,\mathrm{J/K}$, so $\tau = 10^5\,\mathrm{s}$. If the data fix only τ, the sensitivity of a prediction to each parameter is $\partial\hat{y}/\partial R_{\mathrm{th}} = \gamma/R_{\mathrm{th}}$ and $\partial\hat{y}/\partial C_{\mathrm{th}} = \gamma/C_{\mathrm{th}}$, where γ is its sensitivity to $\ln\tau$ and σ the standard deviation of the measurement error, and

$$\mathcal{I}_{\mathrm{raw}} = \frac{\gamma^2}{\sigma^2}\begin{pmatrix} 1/R_{\mathrm{th}}^2 & 1/(R_{\mathrm{th}}C_{\mathrm{th}}) \\ 1/(R_{\mathrm{th}}C_{\mathrm{th}}) & 1/C_{\mathrm{th}}^2 \end{pmatrix} = \frac{\gamma^2}{\sigma^2}\begin{pmatrix} 10^{6} & 10^{-5} \\ 10^{-5} & 10^{-16} \end{pmatrix}.$$

Its large eigenvector is almost exactly the R_{th} axis and its zero eigenvector almost exactly the C_{th} axis, which reads as "the resistance is determined and the capacitance is not". That reading is an artifact of scale: one joule per kelvin is a negligible change of a capacitance of $10^8\,\mathrm{J/K}$, and one kelvin per watt an enormous change of a resistance of $0.001\,\mathrm{K/W}$. The truth is that the product is determined and the split is not, symmetrically.

The remedy is to express each parameter in dimensionless or prior-scaled coordinates,

$$\tilde{\theta}_i = \frac{\theta_i - \mu_i}{\sigma_i}, \tag{15.9}$$

where μ_i and σ_i are the prior mean and standard deviation of parameter i. With prior standard deviations of 50 % of each value, the sensitivity row becomes $\gamma\,(0.5, 0.5)$, the information matrix is proportional to $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, and the eigenvectors are $(1,1)/\sqrt{2}$ and $(1,-1)/\sqrt{2}$, as in the logarithmic example. A flat direction in these coordinates says which parameters are poorly constrained *relative to their plausible ranges*, which is the statement an engineer needs. Logarithmic coordinates are a special case, appropriate when the plausible range of a parameter is proportional to its size.

15.2.6 Structural and practical identifiability

Two questions must be kept apart. *Structural* identifiability asks whether, with perfect noise-free data of the kind the sensors and inputs produce, the parameters could be uniquely recovered. If not, some eigenvalue of the information matrix is exactly zero for every record of that kind, and the problem is fundamental: no quantity of such data helps, and the remedy is to change the model, by reporting the identifiable combination instead of its factors, or to change the measurements, by adding a sensor or an input that cuts across the flat direction. The ratio $\dot{Q}/\dot{m}$ with only ΔT measured, and the product $R_{\mathrm{th}}C_{\mathrm{th}}$ with an unknown heat input, are structural.

Practical identifiability asks whether the actual record pins the parameters down. A model can be structurally identifiable and still be practically unidentifiable because the data are too noisy, too short, or insufficiently exciting: the smallest eigenvalue is positive but small. The red ellipse of Figure 15.4 is practically identifiable with $\lambda_2 = 49.5$; a weaker equilibrium constraint, a standard deviation of 0.5 instead of 0.1 on $\ln R_{\mathrm{th}}$, would leave $\lambda_2 \approx 2$ and a factor-of-two uncertainty along the split direction from a model that is still structurally sound. Here better data do help: a longer record, a larger step, a quieter sensor.

> **Principle 15.2.** *Data determine combinations of parameters. The useful questions are which combinations they determine, read from the eigenvectors of the information matrix in scaled coordinates, and whether a small eigenvalue is structural, fixed by the model and the sensors, or practical, curable with better data of the same kind.*

15.3 Ambiguity in diagnosis

The same idea applies to the causes of a symptom. A diagnosis is identifiable when the competing hypotheses predict different values of something that is measured. This section works through three common ambiguities, the measurements that resolve them, and the cost of adding unknowns to a model.

15.3.1 A rack hot spot

The inlet temperature of one rack rises. Two hypotheses explain it:

- H_1, recirculation: more hot exhaust air reaches the rack inlet;
- H_2, a warmer supply: the cooling unit that serves the rack delivers warmer air.

The mixing rule of Chapter 6, $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{hot}}$, contains both. Take an illustrative baseline with supply air at $T_{\text{supply}} = 18°\text{C}$, hot-aisle air at $T_{\text{hot}} = 36°\text{C}$ and a recirculation fraction $\alpha = 0.1$: the inlet is $0.9 \times 18 + 0.1 \times 36 = 19.8°\text{C}$. The inlet now reads 21.6°C. Under H_1 the supply is unchanged and the recirculation fraction has doubled, $\alpha = (21.6 - 18)/(36 - 18) = 0.2$. Under H_2 the recirculation is unchanged and the supply has warmed to $T_{\text{supply}} = (21.6 - 3.6)/0.9 = 20°\text{C}$. Both reproduce the inlet reading exactly. Without a supply temperature measurement the two hypotheses predict the same inlet temperature, and the inlet sensor cannot distinguish them.

Add a supply-air sensor and they separate: H_1 predicts 18°C at the supply, H_2 predicts 20°C. With a sensor error of $\pm 0.5\,\text{K}$ the two predictions lie four error widths apart, and one reading decides between them. A single sensor has changed the diagnostic capability of the instrumentation from "cannot tell" to "can tell", which no improvement of the inlet sensor could have done. The pattern across racks offers a second, weaker discriminant: a warmer supply raises the inlets of every rack the unit serves, whereas recirculation is usually local.

Table 15.1 collects the signatures of this pair and of the pair in the next subsection. Each row is a hypothesis; each column is a measured or inferred quantity; each entry is the direction in which the hypothesis moves that quantity (↑ up, ↓ down, – unchanged). Two hypotheses are distinguishable by the instrumentation exactly when their rows differ in a column that is actually measured.

Table 15.1: Signatures of four hypotheses on five quantities, illustrative. Hypotheses are distinguishable only through a column that is measured. The airflow is inferred, not measured, from the energy balance.

Hypothesis	T_{in}	T_{supply}	airflow (inferred)	N_{fan} command	Δp_{filter}
H_1 recirculation	↑	–	–	–	–
H_2 warmer supply	↑	↑	–	–	–
H_3 fan degradation	↑	–	↓	↑	↓
H_4 filter blockage	↑	–	↓	↑	↑

15.3.2 Fan degradation and filter blockage

A degraded fan and a blocked filter both lower the airflow delivered by a cooling unit, and with it the airflow the energy balance infers. With only temperatures and powers measured, rows H_3 and H_4 of Table 15.1 agree in every measured column: the inlet warms, the inferred airflow falls, and nothing distinguishes a worn bearing or slipping belt from a clogged filter.

Two further measurements separate them. The fan speed command N_{fan} rises in both cases if a controller is holding a pressure or temperature setpoint, so it confirms that airflow is short but does not by itself say why. The pressure difference across the filter, Δp_{filter}, moves in opposite directions: it rises when the filter is blocked, because the same or a lower flow is forced through a higher resistance, and it falls when the fan is degraded, because less flow passes through an unchanged filter and the pressure drop of a flow resistance scales with the square of the flow, $\Delta p = K\dot{V}^2$. Together they make the diagnosis: a fan commanded high while delivering little pressure across an unchanged filter suggests fan degradation; a high filter pressure difference suggests blockage. Chapter 9 drew this case as a capability map for one rack; here it is one row pair of a signature table.

15.3.3 Sensor bias

A sensor with a constant bias b reads

$$y = T + b. \tag{15.10}$$

With one sensor and nothing else, any pair with $T + b = y$ explains the reading: a reading of 36°C is a true 36°C with no bias, a true 32°C with a bias of +4 K, or a true 30°C with +6 K. The pair (T, b) is structurally non-identifiable; only the sum is identifiable, as in equation (15.6).

Physics can break the ambiguity. Suppose the biased sensor is the rack outlet sensor, and the rack power, the inlet temperature and the airflow are known. The energy balance,

$$T_{\text{out}} = T_{\text{in}} + \frac{P_{\text{rack}}}{\dot{m}_{\text{rack}}\, c_p}, \tag{15.11}$$

then predicts a plausible outlet temperature independently of the sensor, and an outlet reading that violates the balance makes the bias estimable. With an illustrative $T_{\text{in}} = 22$°C, $P_{\text{rack}} = 20$ kW and $\dot{m}_{\text{rack}} = 2\,\text{kg/s} \pm 5\,\%$, the balance puts the true outlet between $22 + 20/2.1 = 31.5$°C and $22 + 20/1.9 = 32.5$°C, and a reading of 36°C implies a bias between 3.5 and 4.5 K.

Only if the physics is informative enough. If the airflow is also unknown, the true temperature, the bias and the airflow are all uncertain, and the same three measurements may not separate them. With the airflow known only to lie between 1.4 and 2.8 kg/s, the balance allows a true outlet from $22 + 20/2.8 = 29.1$°C to $22 + 20/1.4 = 36.3$°C, and the bias anywhere from −0.3 to +6.9 K, a range that includes zero: the data cannot even say whether the sensor is biased.

Figure 15.5 draws the two cases. The horizontal axis is the true outlet temperature, the vertical axis the sensor bias. The diagonal line is the sensor equation (15.10) for the reading $y = 36$°C; every point on it fits the reading. The narrow green band is the range of true outlet temperatures the energy balance allows with the airflow known to 5 %; the wide gold band is the range with the airflow known only to a factor of two. The segment of the diagonal inside each band is the set of explanations left. In the narrow band it is short and lies well above the dotted no-bias line; in the wide band it is long and crosses it.

15.3.4 Latent variables and the cost of complexity

The bias example shows a general effect: each latent variable added to a model is one more unknown competing for the same measurements. If the measurements do not grow with the unknowns, a richer model can be less identifiable than a simpler one. Adding the airflow as an unknown turned an estimable bias into an unestimable one, although the model with the unknown airflow is the more realistic of the two. Richer models are not automatically better.

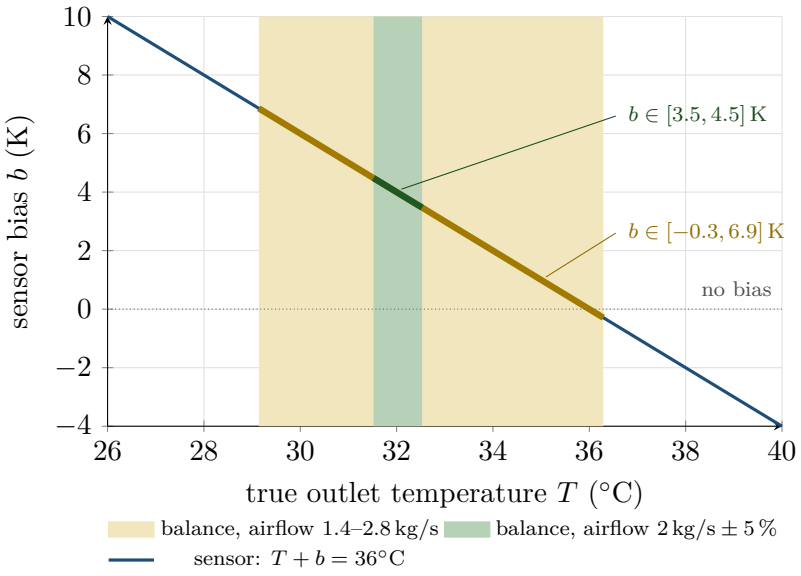


Figure 15.5: Sensor bias and true temperature for one outlet reading of 36°C, $y = T + b$ (equation (15.10)), against the energy balance $T = T_{\text{in}} + P_{\text{rack}}/(\dot{m}_{\text{rack}} c_p)$ (equation (15.11)) with $T_{\text{in}} = 22$°C, $P_{\text{rack}} = 20$ kW, $c_p = 1.0$ kJ/(kg K). A well-known airflow bounds the bias away from zero; a poorly known airflow does not. Values illustrative.

The effect compounds. Compare an illustrative model A with five parameters, all constrained by the data, and a model B with twenty-five parameters, of which fifteen are effectively unconstrained: their information-matrix eigenvalues are close to zero, so along fifteen directions the posterior is the prior. Model B is fragile, because small changes in the data move the poorly constrained parameters a long way; it is hard to interpret, because its estimates mix data and assumption in proportions that differ by parameter; and it is prior-dependent, because three fifths of its answer is whatever was assumed. More physical detail is not more inferential power. The point bears directly on detailed facility models, digital twins and computational fluid dynamics models among them, which Chapter 22 discusses: their detail is useful for prediction under assumed parameters and is not, by itself, evidence that the parameters can be learned from the facility's sensors.

15.4 Priors, adequacy and model structure

Three things can make an analysis look better identified than it is: a prior that supplies the certainty the data lack, a model structure that is wrong but fits, and a choice between structures that the data cannot make. This section treats each and gives the checks that expose them.

15.4.1 Priors that hide non-identifiability

In the Gaussian approximation of Section 15.2.4, the posterior precision is the sum of the data's information and the prior's precision. Where the data contribute nothing, the posterior equals the prior. Suppose the likelihood is nearly flat in R_{th} and the analysis uses a tight prior, $R_{\text{th}} \sim \mathcal{N}(1.0, 0.05^2)$ in K/kW. The posterior then has a standard deviation close to 0.05 K/kW, 5 % of the value, and looks like a well-determined estimate. It is not: the prior identified it, and the measurements did not. Nothing in a reported mean and standard deviation distinguishes the two cases, so an analysis that uses informative priors must say, for each quantity it reports, how much of the certainty came from the prior.

15.4.2 Data-driven and prior-driven certainty

The test is to ask how much the posterior would change if the prior were widened. If it changes dramatically, the data are not informative about that quantity; if it barely changes, they are. The test separates two statements that read alike: "identified by the measurements" and "constrained mainly by prior knowledge".

Figure 15.6 applies it to two illustrative cases. In each panel the horizontal axis is R_{th} in K/kW, enlarged more than five times in the right panel, and the vertical axis is probability density. The gray curve is the likelihood, drawn as a normalized density; the solid red curve is the posterior with the tight prior $\mathcal{N}(1.0, 0.05^2)$; the dashed blue curve is the posterior after the prior is widened tenfold to $\mathcal{N}(1.0, 0.5^2)$. In the left panel the likelihood is nearly flat, a standard deviation of 0.8 K/kW about 1.3. The tight-prior posterior has a standard deviation of 0.050; the wide-prior posterior has 0.42, a factor of 8.5 wider. The data supply $1/0.8^2 = 1.6$ of the 401.6 units of posterior precision, 0.4 %: the estimate is prior-driven. In the right panel the likelihood has a standard deviation of 0.04 about 1.08. The two posteriors have standard deviations of 0.031 and 0.040, a factor of 1.3 (the wide-prior posterior lies almost on the likelihood), and the data supply 625 of the 1025 units of precision under the tight prior, 61 %: the estimate is data-driven.

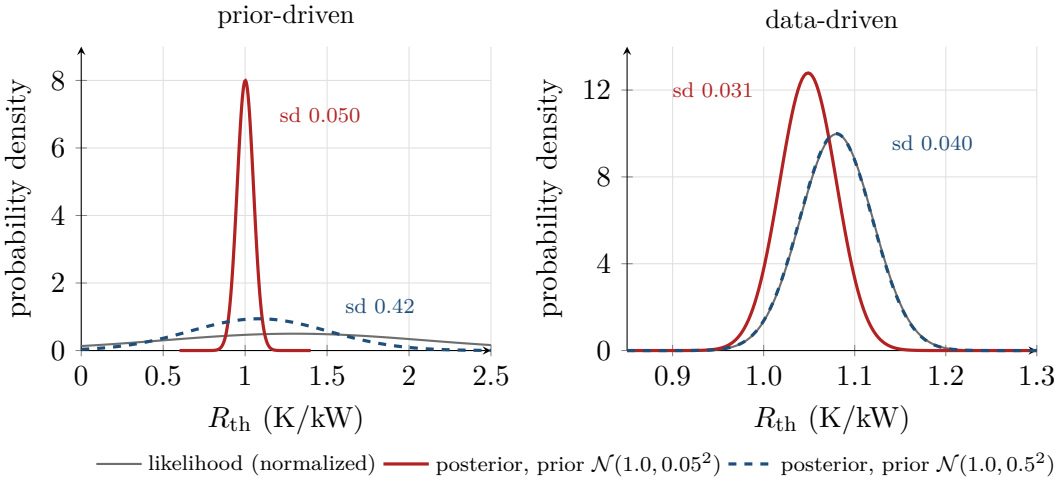


Figure 15.6: The prior-widening test, illustrative. Gaussian likelihood and priors on R_{th}; posterior precision is the sum of likelihood and prior precisions. Left: likelihood $\mathcal{N}(1.3, 0.8^2)$; widening the prior tenfold widens the posterior 8.5 times. Right: likelihood $\mathcal{N}(1.08, 0.04^2)$; the same widening changes the posterior width by a factor of 1.3.

The arithmetic behind the left panel is short. With prior precision $1/0.05^2 = 400$ and data precision $1/0.8^2 = 1.5625$, the posterior precision is 401.56 and its standard deviation $1/\sqrt{401.56} = 0.0499$; its mean is the precision-weighted average $(400 \times 1.0 + 1.5625 \times 1.3)/401.56 = 1.001$. With the wide prior, precision 4, the total is 5.56, the standard deviation 0.424 and the mean 1.084. A reported "$R_{\text{th}} = 1.00 \pm 0.05$ K/kW" from the left panel and one from the right panel look equally authoritative; only the widening test, or a report of the data's share of the precision, tells them apart.

15.4.3 Model adequacy before identifiability

Identifiability is a property of a model, and a model can be wrong. Suppose a rack model omits a major recirculation path. The model then has no mechanism for the extra inlet heating that the path causes, and the fit absorbs it into whatever parameter can produce a similar effect, typically the thermal resistance. With an illustrative rack whose true resistance is 1 K/kW at

8 kW, and 3 K of additional inlet heating from recirculation, a model without recirculation fits a rise of $8 + 3 = 11$ K with $R_{\mathrm{th}} = 11/8 = 1.375$ K/kW. A long, well-excited record may pin that value to 2 %. The estimate is tightly identified and 37.5 % wrong, and its tightness is not evidence of its correctness: it is a precise estimate of a parameter of the wrong model, standing in for a mechanism the model does not contain. The wrong value also predicts wrongly under any change that affects the missing path differently from the resistance, such as a change of blanking panels or of the supply airflow.

Principle 15.3. *Identifiability does not imply correctness. A parameter of an inadequate model can be identified precisely while it absorbs a mechanism the model omits, so adequacy is tested before identifiability is trusted.*

15.4.4 Adequacy, plausibility and identifiability

Three separate questions must therefore be asked of an inference, and Table 15.2 states them. *Adequacy* asks whether the model can reproduce the important observed behavior, judged by the size and structure of its residuals. *Plausibility* asks whether the inferred parameters and states are physically reasonable, judged against equipment ratings, conservation and engineering ranges. *Identifiability* asks whether the data uniquely constrain the quantities of interest, judged by the information matrix and the widening test. A model can pass one and fail the others, and each failure calls for a different response.

Table 15.2: Three tests of an inference, each with its question, its evidence and the typical failure.

Test	Question	Evidence	Typical failure
Adequacy	Can the model reproduce the important observed behavior?	residual size, residual structure in time and by load	a missing mechanism leaves structured residuals
Plausibility	Are the inferred parameters and states physically reasonable?	equipment ratings, conservation, engineering ranges	a parameter absorbs a missing mechanism and leaves its range
Identifiability	Do the data uniquely constrain the quantities of interest?	information-matrix eigenvalues, prior-widening test	a ridge or a prior-driven posterior

A model that fits and is implausible. A reduced cooling model of a room fits the measured temperatures very well, with residuals of a few tenths of a kelvin. Its inferred airflow, however, is ten times what the installed fans can produce: say 40 kg/s against a combined fan rating of about 4 kg/s. Adequacy looks good; plausibility fails. The fit is achieving its small residuals by a physically impossible route, most likely because the airflow is standing in for a heat path or a capacity the model lacks. The correct response is to block every downstream inference that uses this airflow, a capacity estimate, a recirculation fraction, a fan-energy saving, until the model is repaired, however good the fit.

15.4.5 Competing model structures and model evidence

Structural uncertainty can itself be expressed as a probability. A hot spot may be explained by model $\mathcal{M}_A$, recirculation over the top of the rack row, or by model $\mathcal{M}_B$, a local obstruction of

the rack's air path, and both may fit the data similarly. The uncertainty is then at the level of the model, and it is represented by posterior probabilities of the structures,

$$\Pr(\mathcal{M}_j \mid y) \propto p(y \mid \mathcal{M}_j)\,\Pr(\mathcal{M}_j), \qquad p(y \mid \mathcal{M}_j) = \int p(y \mid \theta, \mathcal{M}_j)\, p(\theta \mid \mathcal{M}_j)\,\mathrm{d}\theta, \tag{15.12}$$

where $p(y \mid \mathcal{M}_j)$ is the model evidence: the likelihood averaged over the model's prior, not maximized over it.

The averaging is what keeps a more complex model from being favored automatically. A model with more free parameters spreads its predictive probability over a wider range of possible observations, so it assigns less probability to any particular one. Take an illustrative inlet temperature excess of 3 K. Model $\mathcal{M}_A$, with a single recirculation fraction and a prior range for it, predicts excesses spread evenly between 1 and 5 K, a density of 0.25 per kelvin. Model $\mathcal{M}_B$, with several free obstruction parameters, can produce anything between 0 and 12 K, a density of $1/12 \approx 0.083$ per kelvin. Both fit the observation, but the ratio of evidences is $0.25/0.083 = 3$, and with equal prior probabilities $\Pr(\mathcal{M}_A \mid y) = 0.75$. Had the excess been 8 K, outside what $\mathcal{M}_A$ can produce, its evidence would be zero and $\mathcal{M}_B$ would be certain. The evidence penalizes unnecessary flexibility and rewards it only when the data require it.

15.4.6 Observational equivalence

Model evidence cannot rescue information that is absent. If two models are observationally equivalent under the available sensors, they predict the same distribution of every reading, $p(y \mid \mathcal{M}_A) \approx p(y \mid \mathcal{M}_B)$, and the ratio of evidences is one for every record. The posterior odds then equal the prior odds,

$$\frac{\Pr(\mathcal{M}_A \mid y)}{\Pr(\mathcal{M}_B \mid y)} = \frac{p(y \mid \mathcal{M}_A)}{p(y \mid \mathcal{M}_B)}\,\frac{\Pr(\mathcal{M}_A)}{\Pr(\mathcal{M}_B)} \approx \frac{\Pr(\mathcal{M}_A)}{\Pr(\mathcal{M}_B)},$$

and whatever probabilities the analysis reports are the probabilities it assumed. The conclusion is not "$\mathcal{M}_A$ is more likely" but "the current measurements cannot discriminate between $\mathcal{M}_A$ and $\mathcal{M}_B$". Recirculation over the row and a local obstruction are observationally equivalent for a single inlet sensor at mid-height; they are not for a vertical array of inlet sensors, because recirculation over the top warms the upper inlets first and an obstruction warms the inlets behind it.

15.5 Capability and the limits of inference

The previous sections give the means to recognize when a question exceeds the information in the data. This section states what a disciplined analysis reports in that case and how the capability of a dataset is laid out question by question.

15.5.1 Where the analysis stops

Suppose the question is "is this hot spot caused by recirculation or by reduced airflow?" and the instrumentation cannot separate the two. The right output is not "recirculation, 61 %" unless something in the data supports that number. When the two mechanisms are observationally equivalent under the installed sensors, a probability such as 61 % is a restatement of the prior, and the widening test of Section 15.4.2 would move it at will. The right output is a statement of the evidence: both mechanisms are consistent with the current measurements; a direct airflow measurement at the rack, or a local pressure measurement across its air path, would distinguish them.

A disciplined analysis therefore sometimes stops short of an answer. Its output in that case has two parts: the statement that the requested inference is unsupported by the available

measurements, and the measurement that would make it inferable. That is less impressive than an analysis that always returns a number, and it is more trustworthy, because it makes every number the analysis does return a statement about the data. An unsupported inference identified as such is a result, with an operational use: it directs the next instrumentation decision.

Principle 15.4. *When the measurements cannot separate competing explanations, the correct result is a statement of that fact together with the measurement that would separate them, not a probability supplied by the prior.*

15.5.2 The dataset capability map

The stopping rule applies question by question, and its results can be laid out in advance for a dataset as a capability map: a list of questions, each with the status the instrumentation gives it and the basis for that status. Table 15.3 is an illustrative map for a room instrumented with rack power meters, rack inlet and outlet temperatures at a sparse set of racks, and supply and return temperatures at the cooling units. It distinguishes four grades. Rack heat load is *supported*: it is calculated from a measured power by the relation $\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$. Rack airflow is *supported with uncertainty*: it is inferred from power and temperature difference by equation (15.4), and its uncertainty is that of the temperature difference and of the assumption that all the power leaves in the measured air. The recirculation fraction is *weakly identifiable*: with inlet sensors at a few racks and none at the top of the racks, the mixing rule constrains it only where sensors happen to be. Fan wear against filter blockage is *unsupported*: with neither fan telemetry nor a filter pressure difference, the two hypotheses have the same signature (Table 15.1).

Table 15.3: An illustrative dataset capability map. Each question is graded by what the instrumentation supports, with the basis of the grade and the measurement that would raise it.

Question	Grade	Basis	Would be raised by
Rack heat load	supported	calculated from measured P_{rack}	–
Rack airflow	supported with uncertainty	inferred from P_{rack} and $T_{\text{out}} - T_{\text{in}}$, equation (15.4)	a direct airflow measurement
Recirculation fraction	weakly identifiable	mixing rule at the few racks with inlet sensors	inlet arrays at the racks of interest
Fan wear or filter blockage	unsupported	identical signatures without fan or filter data	N_{fan} and Δp_{filter}

A capability map is written before an analysis is run, from the structure of the model and the list of sensors, and it is revised after the analysis by the information-matrix and widening tests. It turns "can the data answer this?" from a question asked after a confident answer has been given into a property of the dataset that is known beforehand, and it is the natural input to an instrumentation plan.

15.5.3 The value of a measurement depends on the question

A sensor has no value in the abstract. Its value is the change it makes to the answer of a particular question, and two questions about the same room can rank the same sensors in opposite orders.

Consider the question "is the cooling capacity adequate?". A chilled-water flow meter on the loop serving the room is valuable for it: with the supply and return temperatures it gives the heat

the coils actually remove, $\dot{m}_w c_{p,w}(T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}})$, for example $20\,\mathrm{kg/s} \times 4.18\,\mathrm{kJ/(kg\,K)} \times 6\,\mathrm{K} = 502\,\mathrm{kW}$, to compare with the IT load and the plant capacity. A temperature array at one rack adds almost nothing to that question. Now consider "is local recirculation causing the hot spot at rack 12?". A vertical array of inlet temperature sensors on rack 12 is valuable for it: recirculation over the top of the row warms the upper inlets more than the lower, and the array shows that gradient. The chilled-water flow meter adds almost nothing, because the room's total heat removal can be adequate while one rack breathes its own exhaust.

Sensor value is therefore question-dependent, and so is the identifiability a sensor provides. A plan that asks "which sensor is best?" without naming the question has no answer. Chapter 19 turns this into a quantitative rule for choosing the next measurement by the information it adds about a named question.

15.6 Identifiability relative to the decision

Identifiability has so far been judged against the parameters of a model. Operation judges it against a decision. This section shows that a non-identifiable split can be harmless, that a decision can be robust to weakly identified parameters and fragile to well-identified ones, how to attribute the uncertainty of a risk to its sources, and how the chapter's tests combine into one workflow.

15.6.1 An identified net heat balance

Consider a room whose heat balance contains the measured IT power, a constant unmetered heat source P_{const} (lighting, distribution losses, people, solar gain through the envelope) and a constant cooling removal $\dot{Q}_{\mathrm{cool}}$ that is not metered either:

$$C_{\mathrm{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = P_{\mathrm{IT}} + P_{\mathrm{const}} - \dot{Q}_{\mathrm{cool}}. \tag{15.13}$$

The two unmetered terms enter only through their difference $P_{\mathrm{const}} - \dot{Q}_{\mathrm{cool}}$. The data constrain the difference and not the individual values, and the posterior is a ridge along $P_{\mathrm{const}} - \dot{Q}_{\mathrm{cool}} \approx \dot{Q}_{\mathrm{net}}$, a constant the data fix. With an illustrative $P_{\mathrm{IT}} = 480\,\mathrm{kW}$ and a room temperature that holds steady, $\dot{Q}_{\mathrm{net}} = -480\,\mathrm{kW}$, and the pairs $(P_{\mathrm{const}}, \dot{Q}_{\mathrm{cool}}) = (20, 500)$, $(50, 530)$ and $(100, 580)\,\mathrm{kW}$ fit equally. In the scaled coordinates of Section 15.2.5 the information matrix is proportional to $\left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$, with eigenvalues 2 and 0 (times the data precision): the difference direction $(1, -1)/\sqrt{2}$ is determined, the sum direction $(1, 1)/\sqrt{2}$ is flat. The statement the data support is "the net heat balance is identified; the split between these two terms is not".

The operational meaning depends on the question. If the decision needs only the net term, for example the rate at which the room warms or cools at a planned IT load, $\mathrm{d}T/\mathrm{d}t = (P_{\mathrm{IT}} + \dot{Q}_{\mathrm{net}})/C_{\mathrm{th}}$, the ambiguity does not matter: the answer depends on $\dot{Q}_{\mathrm{net}}$, which is known. If the decision is which component to repair, a cooling unit removing less than its rating or an unexpected heat source in the room, the ambiguity matters entirely, because the two repairs correspond to opposite ends of the ridge. Identifiability is judged relative to the decision, not in the abstract.

15.6.2 Decision identifiability

A decision can be identified when its parameters are not. Suppose R_{th} and C_{th} of a zone are individually uncertain, each known only to about 20 %, but every plausible combination predicts a temperature far from the limit. With an illustrative $T_{\mathrm{cool}} = 18\,^{\circ}\mathrm{C}$, a heat input of 8 kW and R_{th} anywhere between 0.8 and 1.2 K/kW, the steady temperature lies between 24.4 and 27.6 °C,

and every combination gives $\Pr(T > 30°\mathrm{C}) < 0.001$. The safety decision, to leave the zone as it is, is robust: the parameters are weakly identifiable and the decision is strongly identifiable.

The reverse also happens near a threshold. Take two models of the same zone, model A predicting a maximum temperature of 29.8°C and model B 30.2°C, against a limit of 30°C. The difference between them may be small in parameter terms: at 12 kW and a coolant at 18°C, it corresponds to thermal resistances of 0.983 and 1.017 K/kW, a 3.4 % difference that the data may not resolve. The decision flips between them: one model says the limit holds, the other that it is exceeded. Here the parameters may be well identified by ordinary standards and the decision still is not.

Figure 15.7 shows the two cases. In both panels the horizontal axis is the predicted maximum temperature in °C and the vertical axis is probability density; the dashed vertical line is the 30°C limit. The left panel is the robust case: the predictive distribution, a normal distribution with mean 26°C and standard deviation 1 K that covers the plausible parameter range, lies almost entirely below the limit, with $\Pr(T > 30°\mathrm{C}) = 3 \times 10^{-5}$. The right panel is the threshold case: models A and B, each with a predictive standard deviation of 0.15 K, straddle the limit, and they assign the exceedance probabilities 0.09 and 0.91.

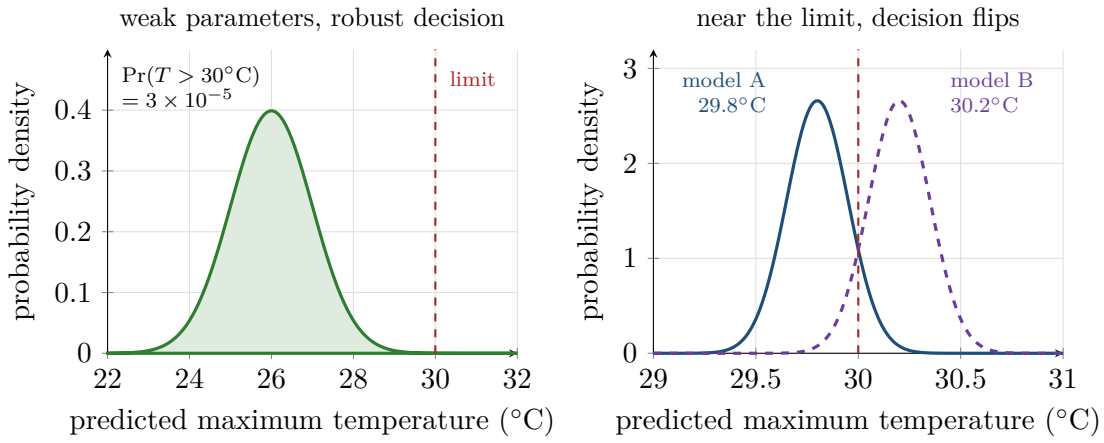


Figure 15.7: Decision identifiability, illustrative. Left: a predictive distribution $\mathcal{N}(26, 1^2)$ in °C spanning the plausible parameter range; the decision is the same everywhere in it. Right: two models with means 29.8 and 30.2°C and standard deviation 0.15 K; exceedance probabilities 0.09 and 0.91 of the 30°C limit (dashed).

15.6.3 Identifiability and risk

The link between the two is a risk metric. Write the risk as the probability that a temperature exceeds its limit,

$$\mathcal{R} = \Pr(T > T_{\max}), \tag{15.14}$$

computed by propagating the posterior over the uncertain parameters and states through the model. The relevant question about each uncertain direction is then how much it moves $\mathcal{R}$. A direction of the posterior that is poorly identified but along which $\mathcal{R}$ barely changes, such as the split between R_{th} and C_{th} for a steady-state limit that depends only on $R_{\mathrm{th}}\,P_{\mathrm{rack}}$, does not need resolving. A direction that is moderately identified but along which $\mathcal{R}$ changes steeply dominates the uncertainty of the risk, and measurement effort belongs there. In the notation of Section 15.2.4, the relevant quantity is the posterior spread along each eigenvector v_i multiplied by the sensitivity of $\mathcal{R}$ along it.

Uncertainty attribution. The same idea can be stated as a budget. The variance of the estimated risk is divided among its sources, for example by the first-order share of each source k, $\mathrm{Var}(\mathbb{E}[\mathcal{R} \mid \theta_k])/\mathrm{Var}(\mathcal{R})$, the fraction of the variance that would disappear if that source alone were known exactly. In an illustrative hot-spot risk the shares are 50 % for the rack airflow, 30 % for the recirculation fraction, 10 % for the rack power and 10 % for sensor noise. The reading is direct: instrument the airflow first. A perfect airflow measurement would remove half of the variance, and the standard deviation of the risk estimate would fall by a factor of $\sqrt{0.5} = 0.71$; a perfect power measurement would remove a tenth of it, a factor of $\sqrt{0.9} = 0.95$. The shares add to one only when the sources do not interact; where they do, the remainder is the interaction share and is reported as such.

Figure 15.8 draws the budget. Each bar is one source of uncertainty, with its share of the variance of the risk estimate on the horizontal axis, in percent. The ordering of the bars is the ordering of the measurement priorities.

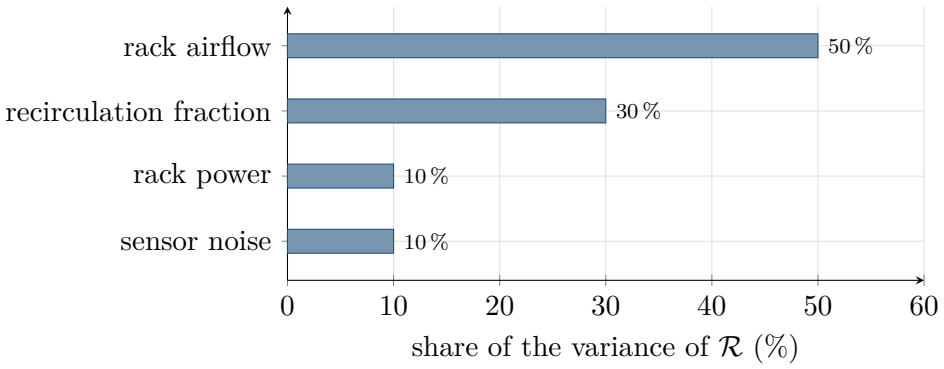


Figure 15.8: Illustrative attribution of the variance of a hot-spot risk estimate $\mathcal{R} = \Pr(T > T_{\max})$ to four sources, by first-order variance share $\mathrm{Var}(\mathbb{E}[\mathcal{R} \mid \theta_k])/\mathrm{Var}(\mathcal{R})$. The largest share names the first measurement to add.

15.6.4 The inference workflow

The tests of the chapter combine into a sequence of eight steps, which Figure 15.9 draws. The steps run top to bottom in the left column; the boxes on the right are the exits that a failed test leads to, and the three boxes at the bottom are the three possible outputs.

1. Define the target quantity, in terms of the decision it serves.
2. Identify the physical equations that constrain it: balances, component relations, the mixing rule.
3. Map the available measurements onto those equations: which are measured, which calculated, which must be inferred.
4. Test model adequacy: residual size and structure.
5. Test identifiability: the information matrix in scaled coordinates, the prior-widening test, plausibility of the estimates.
6. Quantify the posterior uncertainty of the target.
7. Decide whether the inference is decision-useful: whether the decision is the same across the posterior.
8. If it is not, recommend the new measurement that would make it so, chosen by attribution.

The workflow has two properties worth stating. Adequacy is tested before identifiability, for the reason of Section 15.4.3: an identifiability analysis of an inadequate model returns confident statements about the wrong parameters. And the last test is against the decision, for the reason of Section 15.6.2: a posterior can be wide and decision-useful, or narrow and not.

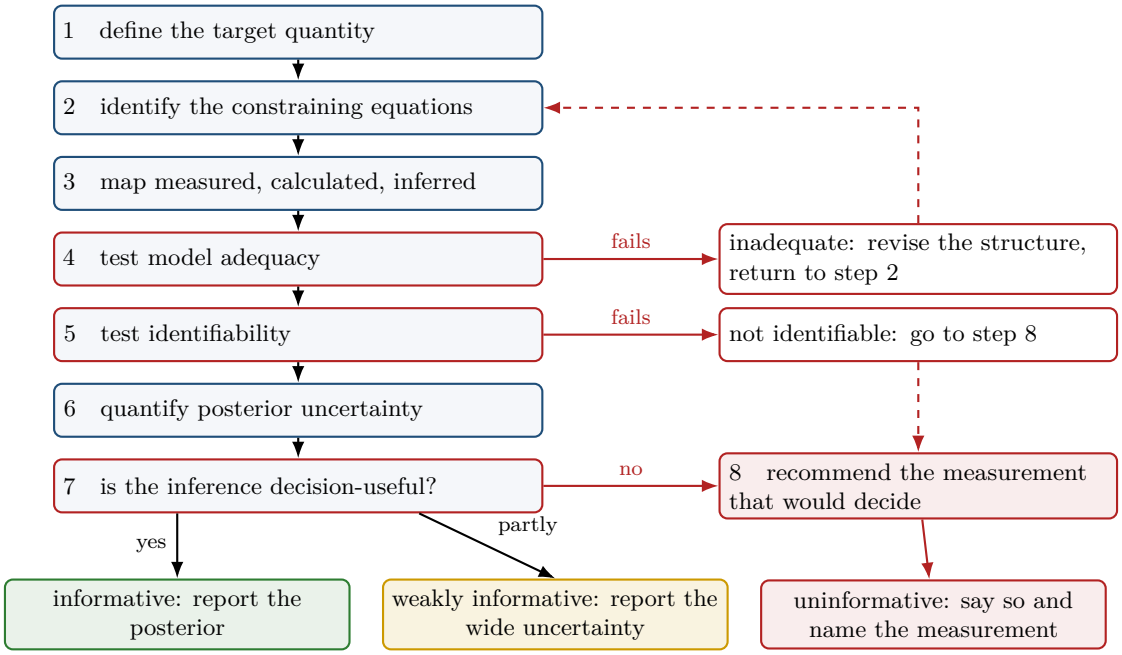


Figure 15.9: The eight-step inference workflow. Steps 4, 5 and 7 (red outline) are tests; a failed adequacy test returns to the model structure, a failed identifiability or usefulness test leads to a measurement recommendation. The three outputs at the bottom correspond to informative, weakly informative and uninformative data.

15.6.5 Four distinctions

The chapter rests on four words that are often used interchangeably and mean different things. Table 15.4 lists them. *Noise* is a property of a sensor: it is imprecise. *Uncertainty* is a property of knowledge: the state or parameter is not known exactly. *Non-identifiability* is a property of the model and the measurement set together: several states or parameter combinations explain the same evidence. *Model inadequacy* is a property of the model alone: it does not represent reality well enough. Each has its own remedy, and applying the remedy of one to another, more data against non-identifiability, a tighter prior against inadequacy, produces confidence without knowledge.

Table 15.4: Four distinctions, with what each is a property of, what reduces it, and an example from the chapter.

Distinction	Property of	Reduced by	Example
Noise	the sensor	a better sensor, averaging	$T = 24.0 \pm 0.5$°C
Uncertainty	the state of knowledge	more and better data	$p(R_{\text{th}} \mid y)$ narrowing with the record
Non-identifiability	the model with its sensors and inputs	an independent measurement or input	$\dot{Q}$ and $\dot{m}$ from ΔT alone; R_{th} and C_{th} from τ
Model inadequacy	the model structure	a better structure	a missing recirculation path absorbed into R_{th}

The deeper point is that not every unknown quantity is inferable from every dataset, and no inference machinery changes that. The physics of the facility and the geometry of its

measurements determine what information exists. The question to ask of a dataset is whether it contains information capable of distinguishing the competing explanations. If it does, the posterior is estimated and reported. If it does so only weakly, the wide uncertainty is reported as the result. If it does not, the analysis says so and names the measurement that would change the situation. That restraint is a source of credibility, not a lack of capability.

15.7 What to retain

Definition 15.1 (Identifiability)**.** A quantity is identifiable from a dataset when no two different values of it explain the measurements equally well. Uncertainty shrinks with more and better data of the same kind; non-identifiability yields only to independent information, and is judged relative to the decision the inference serves: if the data inform the decision, estimate it; if weakly, report the wide uncertainty; if not, say so and name the measurement that would.

Table 15.5 lists the chapter's objects and the subsection that defines each.

Table 15.5: The objects of uncertainty and identifiability and where each is defined.

Object	Expression	Role	Defined in
Four kinds of uncertainty	measurement, parameter, state, structural	where uncertainty enters	15.1.1
Identifiability	equation (15.3)	uniqueness of explanation	15.1.3
Independent constraint	$\dot{m} = P_{\text{rack}}/(c_p \Delta T)$	cuts across a flat direction	15.1.4
Excitation	varied $P_{\text{rack}}(t)$ in equation (15.5)	reveals C_{th}	15.2.1
Identifiable combination	$\phi(\theta)$, equation (15.6)	what to report	15.2.2
Correlation	ρ, equation (15.7)	orientation, not size	15.2.3
Fisher information	$\mathcal{I} = J^{\mathsf{T}}\Sigma_{\epsilon}^{-1}J$, eigenpairs λ_i, v_i	flat directions	15.2.4
Scaled coordinates	$\tilde{\theta}_i = (\theta_i - \mu_i)/\sigma_i$	unit-free flat directions	15.2.5
Structural, practical	zero or small eigenvalue	fundamental or curable	15.2.6
Signature table	Table 15.1	which sensor separates which causes	15.3.1
Prior-widening test	posterior width under a wider prior	data- or prior-driven	15.4.2
Three tests	adequacy, plausibility, identifiability	gates on an inference	15.4.4
Model evidence	$p(y \mid \mathcal{M}_j)$, equation (15.12)	structural uncertainty	15.4.5
Capability map	Table 15.3	what a dataset supports	15.5.2
Decision identifiability	same decision across the posterior	robustness of an action	15.6.2
Uncertainty attribution	$\text{Var}(\mathbb{E}[\mathcal{R} \mid \theta_k])/\text{Var}(\mathcal{R})$	the next measurement	15.6.3

15.8 Exercises

Exercise 15.1. A rack shows a temperature rise of $\Delta T = 12 \pm 0.5\,\mathrm{K}$, with $c_p = 1.0\,\mathrm{kJ/(kg\,K)}$. (a) List three pairs $(\dot{Q}, \dot{m})$ consistent with an exact 12 K and state what the reading identifies. (b) The rack power meter reads $P_{\mathrm{rack}} = 30 \pm 1.5\,\mathrm{kW}$. Compute the central airflow and its range. (c) The reading is repeated 5000 times while nothing changes. By how much does the uncertainty of ΔT fall, and does the range of (b) change if the power meter is not read again? Explain in terms of independent constraints.

Exercise 15.2. In scaled coordinates $(\tilde{R_{\mathrm{th}}}, \tilde{C_{\mathrm{th}}})$ a record gives the information matrix $\mathcal{I} = \left(\begin{smallmatrix} 900 & 800 \\ 800 & 800 \end{smallmatrix}\right)$. (a) Compute its eigenvalues and condition number. (b) Compute the posterior covariance $\mathcal{I}^{-1}$, the two standard deviations and the correlation. (c) Is the pair identifiable, and is the remaining problem structural or practical? (d) Which eigenvalue would be zero if the record contained no equilibrium information, and what would then be identifiable?

Exercise 15.3. A parameter has a Gaussian likelihood $\mathcal{N}(2.0, 0.6^2)$. (a) Compute the posterior mean and standard deviation under the prior $\mathcal{N}(1.5, 0.1^2)$ and under the widened prior $\mathcal{N}(1.5, 1.0^2)$. (b) Compute the ratio of the two posterior standard deviations and the data's share of the posterior precision in each case. (c) Is a report of the first posterior a statement about the measurements? (d) What likelihood standard deviation would make the data supply half the posterior precision under the tight prior?

Exercise 15.4. A zone has a limit of 30°C. (a) Model A predicts a maximum temperature distributed as $\mathcal{N}(28.5, 0.8^2)$ in °C; compute $\Pr(T > 30°\mathrm{C})$ and say whether the decision "no action" holds for a tolerance of 0.01. (b) Model B predicts $\mathcal{N}(29.6, 0.3^2)$; repeat. (c) The variance of the risk estimate is attributed 50 % to airflow, 30 % to recirculation, 10 % to rack power and 10 % to sensor noise. An airflow meter removes 80 % of the airflow share. By what factor does the standard deviation of the risk estimate fall? (d) Two structures $\mathcal{M}_A$ and $\mathcal{M}_B$ predict the observed excess with densities 0.2 and 0.05 per kelvin. With prior probabilities 0.4 and 0.6, compute $\Pr(\mathcal{M}_A \mid y)$.

15.9 Connection to the TERRA framework

In TERRA a data center is a set of coupled conservation networks, electrical power, air and water mass and thermal energy, with measurements attached to some of their nodes and edges. Identifiability is a property of that graph together with its measurement set, and the chapter's objects map onto it directly.

- **Each balance is a constraint, and identifiability counts them.** The rack node carries the electrical edge P_{rack} (measured), the air edge $\dot{m}_{\mathrm{rack}}$ (inferred) and the heat balance $\dot{Q} = \dot{m} c_p \Delta T$ between the inlet and outlet temperature nodes (measured). Whether $\dot{m}_{\mathrm{rack}}$ is identifiable is the question of Section 15.1.4: does an independent edge, here the power meter, cut across the flat direction that ΔT leaves.
- **The posterior reports its shape, not only its widths.** The Laplace posterior of the inference has the information matrix of Section 15.2.4 as its precision. Read in prior-scaled coordinates, its small eigenvalues and their eigenvectors name the flat directions, the R_{th}–C_{th} split of a thermal node or the P_{const}–$\dot{Q}_{\mathrm{cool}}$ split of a room balance, and the identifiable combination is what gets reported.
- **Each inferred quantity is tagged data-driven or prior-driven.** The prior-widening test of Section 15.4.2 and the data's share of the posterior precision accompany every reported parameter, so that a coil conductance or a recirculation fraction identified by the record is not presented in the same way as one held by its prior.

- **Structural uncertainty is a set of candidate graphs.** A recirculation edge from the hot aisle to a rack inlet ($\mathcal{M}_A$) and a raised resistance on the rack's air edge ($\mathcal{M}_B$) are two graphs scored by model evidence, equation (15.12), after the adequacy and plausibility tests of Table 15.2; when they are observationally equivalent the result says so.
- **TERRA should sometimes stop.** For a query whose answer the measurement set cannot separate, the correct output is the statement that the inference is unsupported and the sensor that would support it; TERRA should remain silent on the split rather than report a prior as a probability. The capability map of Table 15.3 is this rule applied to every query in advance.
- **Risk and attribution choose the next measurement.** The risk $\mathcal{R} = \Pr(T > T_{\max})$ is the forward simulation of the graph from the posterior; its variance budget, Figure 15.8, names the edge or node whose measurement would most reduce it.

In summary, uncertainty is a width that more and better data narrow, while non-identifiability is a set of explanations that fit the evidence equally and that only independent information separates. Identifiability appears as flat directions of the information matrix, can be hidden by priors and by inadequate models, extends to competing model structures, and is ultimately judged against the decision: an analysis estimates what the data inform, reports wide uncertainty where they inform weakly, and states what it cannot infer together with the measurement that would change that. The next chapter, Chapter 16, uses these tools to rank competing fault hypotheses by their posterior probability given the measurements.

Chapter 16

Probabilistic Diagnosis

16.1 Diagnosis as inference

Earlier chapters listed the ways a data center fails, the residuals that expose a failure, the estimators that reconstruct hidden states and the limits on what a set of measurements can determine. This chapter combines them into one procedure for the moment an abnormal reading appears. The question is

given abnormal measurements, how are competing explanations ranked while accounting for measurement uncertainty, for the prior reliability of each component and for physical consistency?

The central quantity is the probability of each hypothesis H_i given the measurements y,

$$\Pr(H_i \mid y),$$

and the chapter builds it, explains where each of its factors comes from, shows where it stops being informative, and turns it into an action.

Jargon: hypothesis, prior, likelihood, posterior, evidence, marginalization, Bayes factor, filtering, smoothing, anomaly score, entropy, value of information, active diagnosis

A *hypothesis* H_i is a candidate explanation of the measurements, such as "rack 12 receives recirculated hot air." Its *prior* probability $\Pr(H_i)$ is its plausibility before the present measurements are considered. The *likelihood* $\Pr(y \mid H_i)$ is the probability, or probability density, of the measurements actually observed if H_i were true. The *posterior* $\Pr(H_i \mid y)$ is the probability of the hypothesis after the measurements are taken into account. The *evidence* or *marginal likelihood* of a hypothesis is its likelihood averaged over everything the hypothesis leaves unspecified, such as unmeasured states and uncertain parameters; that averaging is called *marginalization*. A *Bayes factor* is the ratio of the likelihoods of two hypotheses for the same measurements. *Filtering* estimates the present state from measurements up to the present; *smoothing* estimates a past state from measurements taken before and after it. An *anomaly score* is a number that grows as a measurement moves away from what is usual, without naming a cause. The *entropy* of a probability distribution measures how spread out it is, in bits; the *value of information* of a measurement is the expected reduction in that entropy, or in expected loss, that the measurement would bring. *Active diagnosis* deliberately perturbs the system, within safe limits, so that the hypotheses predict different responses.

Key idea

Diagnosis is not the assignment of one fault label. It is the maintenance of a physically consistent probability distribution over competing explanations, $\Pr(H_i \mid y) \propto \Pr(y \mid H_i)\ \Pr(H_i)$, in which the likelihood comes from a physical model of what each explanation would make the sensors read, and the distribution is used to rank risks and to choose the next action or the next measurement.

16.1.1 Diagnostic practice and explicit inference

The approach of this chapter goes beyond the usual operating curriculum of a data center. That curriculum follows a fixed sequence: an alarm fires, a fault is detected, a technician troubleshoots, and a maintenance action follows. The tools of each step are rules, alarm thresholds, troubleshooting checklists and, increasingly, anomaly detectors trained on historical telemetry. The operational problem behind the sequence is real and common: a high rack temperature, a cooling unit that underperforms, two sensors that disagree, airflow that is insufficient for no visible reason, a valve or actuator that does not do what it is commanded to, equipment whose performance declines slowly, and failures that interact. What these symptoms share is ambiguity: each is compatible with several causes, and the checklist does not say how strongly the available readings favor one cause over another.

The chapters before this one are mostly fundamentals: how power, heat, air and water behave, and what the sensors and controls report about them. This chapter and the neighboring ones treat the same operating problems as explicit probabilistic inference on a physical model. The difference from the alarm sequence is not that rules or checklists are wrong, but that every step of the inference is stated: which causes are considered, how plausible each was beforehand, what each predicts for every available sensor, and how far the readings move the balance between them.

16.1.2 The posterior over hypotheses

The output of probabilistic diagnosis is a list of hypotheses with a probability attached to each, not a single verdict. The list is always conditional on two things that must be stated with it: the measurements used and the model assumptions under which the likelihoods were computed. A sound statement of a result reads, for example, "given the current evidence and model assumptions, recirculation is the leading explanation," and not "the fault is recirculation."

The book's distinction between measured, calculated and inferred quantities applies directly. A rack inlet temperature and a rack electrical power are measured. The heat a rack releases is calculated from its electrical power. A recirculation fraction, a rack airflow and the probability that a given fault is present are inferred: none is read from an instrument, and each is a conclusion that depends on a model and on the measurements fed to it. The posterior is the most inferred quantity of all, and the rest of the chapter is about making that inference explicit enough to be checked.

16.2 Bayes' rule on a worked case

This section states Bayes' rule, sets up one diagnostic case, a rack whose inlet temperature has risen, and works the posterior through with explicit numbers for every prior and likelihood.

16.2.1 Bayes' rule

For a set of mutually exclusive hypotheses $H_1, \ldots, H_n$, one of which is assumed true, Bayes' rule gives the posterior of each as

$$\Pr(H_i \mid y) = \frac{\Pr(y \mid H_i)\ \Pr(H_i)}{\Pr(y)}, \qquad \Pr(y) = \sum_{j=1}^{n} \Pr(y \mid H_j)\ \Pr(H_j). \tag{16.1}$$

The denominator $\Pr(y)$ is the same for every hypothesis; it only makes the posteriors sum to one. The ranking is therefore set by the numerator,

$$\Pr(H_i \mid y) \propto \Pr(y \mid H_i)\ \Pr(H_i), \tag{16.2}$$

which multiplies two factors with different origins. The prior $\Pr(H_i)$ measures how plausible the hypothesis was before the present data: how often that fault occurs, how old the equipment is, what maintenance has been done. The likelihood $\Pr(y \mid H_i)$ measures how well the data agree with what the hypothesis predicts. A hypothesis can rank high because it is common, because it explains the data well, or both; it ranks low if it is rare and explains the data no better than its competitors, or if it predicts something the data contradict.

16.2.2 The symptom and the candidate explanations

Consider an illustrative case. The inlet temperature of rack 12 reads 29°C, against 23°C in normal operation. Five explanations are considered:

- H_1, high IT load: the rack's servers draw more power and release more heat;
- H_2, reduced airflow: the cold air delivered to the rack's position falls short of what its servers draw, so the difference is drawn from warmer air;
- H_3, hot-air recirculation: exhaust from the hot aisle returns to the rack inlet through a local path, such as a missing blanking panel or an open containment door;
- H_4, warm supply air: the computer room air handler (CRAH) serving the aisle delivers air warmer than its setpoint;
- H_5, sensor bias: the inlet sensor reads high and the air is in fact normal.

The inlet reading alone cannot distinguish them, because every one of them was chosen precisely because it can produce a high inlet reading. What separates them is what each predicts for the other sensors.

16.2.3 Additional evidence

Four more readings are available, all measured: the rack electrical power P_{rack} is normal; the CRAH supply temperature T_{supply} is normal; the hot-aisle temperature T_{hot} is high; and the differential pressure Δp between the cold aisle and the hot aisle is low. Table 16.1 lists, for each hypothesis, what it predicts for each of the five readings. The rows are the hypotheses; the columns are the readings; an entry says whether the hypothesis predicts the reading to rise, to stay normal, or leaves it unaffected, in which case the reading behaves as it does on an ordinary day.

Read against the observed row, the table already says most of what the numbers will say. High IT load predicts a rise in P_{rack}, which is not observed, so $\Pr(y \mid H_1)$ falls. Warm supply air predicts a rise in T_{supply}, which is not observed. Sensor bias predicts that nothing else changes, yet the hot aisle is high. Recirculation predicts a high hot aisle and possibly a low pressure difference, and both are observed. Reduced airflow predicts the same directions, with a weaker link to the pressure difference for the reason given in Section 16.3.1.

Table 16.1: Predicted signatures of the five hypotheses for rack 12. "Up" and "down" are the predicted direction; "normal" means the hypothesis predicts no change; "unaffected" means the hypothesis has no bearing on that reading. The last row is what was observed.

Hypothesis	T_{in}	P_{rack}	T_{supply}	T_{hot}	Δp
H_1 high IT load	up	up	normal	up	unaffected
H_2 reduced airflow	up	normal	normal	up	down, weakly
H_3 recirculation	up	normal	normal	up	down, possibly
H_4 warm supply	up	normal	up	up	unaffected
H_5 sensor bias	up (reading only)	normal	normal	unaffected	unaffected
Observed	29°C	normal	normal	high	low

16.2.4 The posterior ranking

To turn the signatures into numbers, each hypothesis is given an illustrative prior and, for each reading, an illustrative probability of observing what was observed. Where a hypothesis has no bearing on a reading, the probability is the base rate of that observation on an ordinary day: 0.3 for a low pressure difference and 0.2 for a high hot-aisle temperature. The five readings are taken as conditionally independent given the hypothesis, meaning that once the hypothesis is fixed, the noise on one sensor carries no information about the noise on another; the likelihood of all five readings is then the product of the five factors,

$$\Pr(y \mid H_i) = \prod_{k=1}^{5} \Pr(y_k \mid H_i). \tag{16.3}$$

Table 16.2 carries out the computation. The first numeric column is the prior. The next five are the factors of equation (16.3); for example, 0.05 under H_1 in the P_{rack} column is the probability that the power meter reads normal when the IT load has in fact risen, which requires a meter error large enough to hide the rise. The column $\Pr(y \mid H_i)$ is the product of the five factors, the next column is prior times likelihood, and the last column divides each entry by their sum.

Table 16.2: Bayes' rule for rack 12 with illustrative priors and likelihood factors. Each factor is the probability of the observed reading under the hypothesis; their product is the likelihood, and the posterior is prior times likelihood divided by $\Pr(y) = 0.050112$.

Hypothesis	prior	T_{in}	P_{rack}	T_{supply}	T_{hot}	Δp	$\Pr(y \mid H_i)$	prior×lik.	posterior
H_1 high IT load	0.30	0.6	0.05	0.9	0.8	0.3	0.00648	0.001944	0.039
H_2 reduced airflow	0.10	0.6	0.9	0.9	0.8	0.4	0.15552	0.015552	0.310
H_3 recirculation	0.10	0.6	0.9	0.9	0.9	0.6	0.26244	0.026244	0.524
H_4 warm supply	0.10	0.5	0.9	0.05	0.8	0.3	0.00540	0.000540	0.011
H_5 sensor bias	0.40	0.3	0.9	0.9	0.2	0.3	0.01458	0.005832	0.116
Sum	1.00							0.050112	1.000

The arithmetic for recirculation is

$$\Pr(y \mid H_3) = 0.6 \times 0.9 \times 0.9 \times 0.9 \times 0.6 = 0.26244,$$
$$\Pr(H_3 \mid y) = \frac{0.10 \times 0.26244}{0.050112} = \frac{0.026244}{0.050112} = 0.524,$$

and the other rows follow the same pattern. Rounded, the posterior ranking is recirculation 52 %, reduced airflow 31 %, sensor bias 12 %, high IT load 4 %, warm supply 1 %.

Figure 16.1 shows how the evidence moved the distribution. The horizontal axis lists the five hypotheses; for each, three bars give its probability in percent: the prior (light), the posterior after the inlet reading alone (medium), and the posterior after all five readings (dark). After the inlet reading alone, high IT load leads with 38 %, because it had the largest prior among the physical causes and the inlet factor barely discriminates; sensor bias follows with 26 %. The remaining four readings reverse the order: the normal rack power removes most of the probability of high IT load, the normal supply temperature removes warm supply air, the high hot aisle removes most of sensor bias, and the low pressure difference favors recirculation over reduced airflow. Recirculation started with a prior of 10 %, a quarter of that of sensor bias, and finished first.

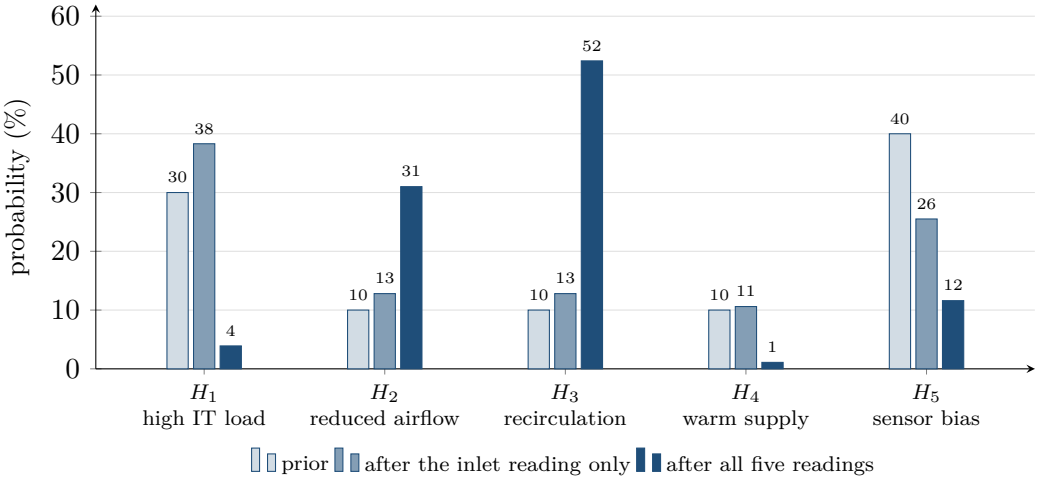


Figure 16.1: Prior and posterior probabilities of the five hypotheses for rack 12, computed by equation (16.1) with the illustrative priors and likelihood factors of Table 16.2. The inlet reading alone leaves high IT load in first place; the four further readings move most of the probability to recirculation and reduced airflow.

Principle 16.1. *A posterior ranking is set by prior times likelihood. Evidence that a hypothesis predicts wrongly, such as a normal rack power under high IT load, removes it more decisively than evidence it predicts correctly supports it, because the likelihood factor of a contradicted prediction is small.*

16.3 Likelihoods from physical models

The likelihood factors of Table 16.2 were stated, not derived. This section shows where they come from in a physical model, how the priors are set and how they change, and how unmeasured states and uncertain parameters enter the likelihood through marginalization, which also penalizes explanations that are too flexible.

16.3.1 Where the physics enters

Each hypothesis imposes a different structure on the measurements, and the structure comes from the relations of the earlier chapters.

Recirculation. The inlet air of a rack is a mixture of supply air and hot exhaust, with recirculation fraction α (Chapter 6):

$$T_{\text{in}} = (1 - \alpha)\, T_{\text{supply}} + \alpha\, T_{\text{hot}}. \tag{16.4}$$

With an illustrative supply of $T_{\text{supply}} = 20°\text{C}$ and a normal hot aisle at $T_{\text{hot}} = 35°\text{C}$, a normal inlet of 23°C corresponds to $\alpha = (23 - 20)/(35 - 20) = 0.2$. Under H_3 the local path carries more exhaust: if the hot aisle near rack 12 rises to 38°C, an inlet of 29°C requires $\alpha = (29 - 20)/(38 - 20) = 0.5$. The hypothesis therefore predicts a normal T_{supply}, a raised T_{hot} and, when the path is an opening between the aisles, a reduced pressure difference across it, which is the pattern observed.

Reduced airflow. The heat a rack releases is carried away by its air, $\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, (T_{\text{out}} - T_{\text{in}})$ (Chapter 5). An illustrative rack at 20 kW with $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$ and a temperature rise of 12 K draws $\dot{m}_{\text{air}} = 20/(1.0 \times 12) = 1.67\,\text{kg/s}$. If the cold air delivered at its position falls short of that, the servers draw the difference from the hot aisle and the inlet warms; if the rack's own airflow falls to 1.25 kg/s, its temperature rise grows to $20/1.25 = 16\,\text{K}$ and its exhaust warms. Either way the hot aisle rises. The link to the measured pressure difference is weaker than for recirculation: where the CRAH fans are controlled to hold the cold-aisle pressure (Chapter 10), a local supply shortfall at one tile need not move the aisle reading.

High IT load. The heat a rack releases equals its electrical power to within a few percent, $\dot{Q}_{\text{rack}} \approx P_{\text{rack}}$ (Chapter 5). A hypothesis of more heat is therefore a hypothesis of more measured power. If P_{rack} is metered with an error of about 0.5 kW and H_1 requires a rise of several kilowatts to explain a 6 K inlet rise, a normal power reading lies many standard deviations from the prediction, which is why its factor in Table 16.2 is 0.05 and not 0.5.

The likelihood of each hypothesis is thus a forward calculation: assume the hypothesis, compute the readings the physics predicts, compare them with the readings observed, and convert the mismatch into a probability using the sensors' error distributions.

16.3.2 Physical reasoning and pattern matching

A classifier trained on labeled history learns a direct mapping from measurements to a fault label, $y \to H$. A physics-constrained diagnosis reasons in the causal direction, from fault to physical state to measurements, $H \to x \to y$, and then inverts the chain with Bayes' rule to obtain $\Pr(H, x \mid y)$.

Figure 16.2 draws the two. The upper row is the classifier: one box for the measurements, one for the label, and a single learned arrow between them. The lower row is the physical model: the hypothesis (here recirculation) sets a physical state (a recirculation fraction and a hot-aisle temperature), and the state sets the measurements through the sensor relations; each solid arrow is labeled with the probability it contributes. The dashed arrow running back along the bottom is the inference.

The difference matters in three ways. The generative model explains its conclusion: the leading hypothesis comes with the physical state that would produce the readings, here $\alpha \approx 0.5$, which an engineer can check. It does not need labeled examples of every fault, which a data center rarely has, because the forward physics supplies the predictions. And it extends to readings never seen in training: a new sensor enters the chain as one more arrow from x to y, and a classifier would have to be retrained.

16.3.3 Priors and their sources

The prior is the probability a hypothesis carries before the present measurements. It should differ between hypotheses, because faults differ in how often they occur. A temperature sensor

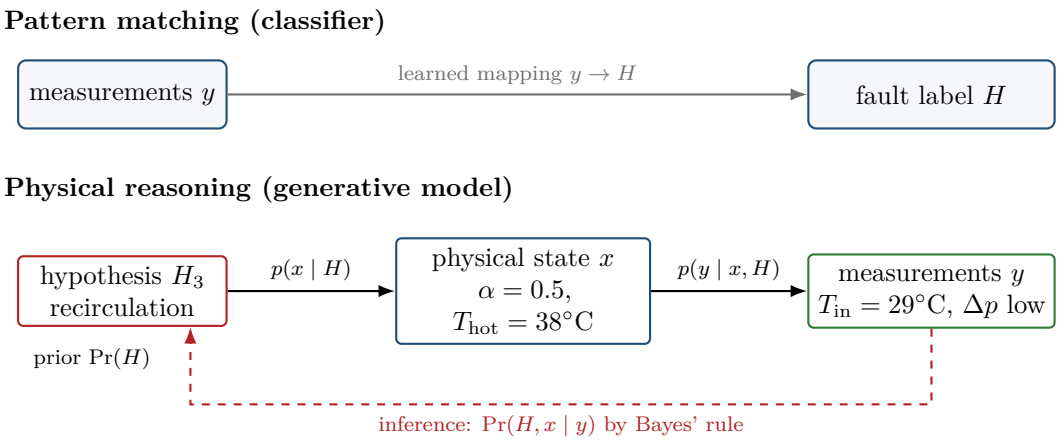


Figure 16.2: A classifier maps measurements to a label directly (top). A physical model generates measurements from a hypothesis through a physical state (bottom, solid arrows), and diagnosis inverts that chain (dashed arrow). The intermediate state is what makes the conclusion interpretable: the posterior names a recirculation fraction, not only a label.

that has drifted is common; a sudden failure of a chiller is rare; the rarer event receives the lower prior. In Table 16.2 sensor bias received the largest prior, 0.40, for that reason.

Priors have identifiable sources: maintenance history (when a sensor was last calibrated, when a filter was last replaced), failure rates of the component class (Chapter 12 develops them), equipment age, operating condition (a unit running near its rated capacity, or through a heat wave), and recent alarms on related equipment. Each is a statement about the facility before the present readings, which is exactly what a prior is.

Priors change over time. A prior is not a constant. The probability of a blocked filter grows with the time since the filter was last replaced; the probability of a chiller failure is lower the day after the chiller was serviced than a year later. A diagnosis system that keeps its priors current, from the maintenance log and the age of each component, ranks the same symptom differently on different days, and it should.

16.3.4 Hidden states and uncertain parameters

A hypothesis rarely fixes everything. H_2 says the airflow at rack 12 is reduced, but not by how much; the deficit is a hidden state x. The likelihood of the hypothesis is then the likelihood averaged over every state the hypothesis allows,

$$\Pr(y \mid H_2) = \int \Pr(y \mid x, H_2)\, p(x \mid H_2)\, \mathrm{d}x. \tag{16.5}$$

The integral asks whether the hypothesis has any plausible configuration of states that explains the observations, and how much of its plausible range does so.

An illustrative calculation shows what the integral does. Let the deficit δ be the fraction of the rack's air demand not met by cold supply, so that the inlet reads $T_{\text{in}} = 23 + 15\,\delta$ in °C (each unit of deficit replaced by air 15 K warmer). Under H_2 let δ be equally likely anywhere between 0 and 0.6, and let the inlet sensor have a Gaussian error with a standard deviation of 0.5 K. The reading 29°C is matched by $\delta = 0.4$, and the deficits that match it within the sensor's error lie in a band of width about $0.5/15 = 0.033$ on either side. The integral evaluates to

$$\Pr(y \mid H_2) = \frac{1}{0.6} \times \frac{1}{15} = 0.111 \text{ per kelvin},$$

while the best single deficit, $\delta = 0.4$, gives the peak sensor density $1/(0.5\sqrt{2\pi}) = 0.798$ per kelvin. The ratio, $0.111/0.798 = 0.139$, is the fraction of the hypothesis's prior range that is consistent with the data, $\sqrt{2\pi} \times 0.033/0.6$.

Parameter uncertainty. Parameters such as a coil's conductance or a fan's curve are uncertain as well, and the full likelihood averages over them too,

$$\Pr(y \mid H_i) = \iint \Pr(y \mid x, \theta, H_i)\, p(x, \theta \mid H_i)\, \mathrm{d}x\, \mathrm{d}\theta. \tag{16.6}$$

With tens of states and parameters the integral has tens of dimensions. It is computationally demanding, and in practice it is approximated: by a Gaussian around the most probable state and parameter values, by sampling, or by exploiting the structure of the model, which Section 16.4.3 takes up.

16.3.5 Evidence and flexibility

The marginal likelihood of equation (16.6) is called the evidence for the hypothesis, and it behaves differently from the best-fit residual. A model with twenty adjustable parameters always fits better than a model with two: with enough parameters any residual can be driven down. The evidence does not reward that. It averages the fit over the whole prior range of the parameters, so a model whose good fit is confined to a small, finely tuned region of a large parameter space receives a small evidence. Approximately,

$$\ln \Pr(y \mid H_i) \approx \ln \Pr(y \mid \hat{x}, \hat{\theta}, H_i) + \sum_k \ln \frac{\sqrt{2\pi}\, \sigma_{\mathrm{post},k}}{\Delta_k}, \tag{16.7}$$

where $\hat{x}, \hat{\theta}$ are the best-fit values, and for each adjustable quantity k, $\sigma_{\mathrm{post},k}$ is the width of the range the data allow and Δ_k the width of the range the hypothesis allowed beforehand. The ratio is the 0.139 of the previous subsection, generalized to many quantities, and each term of the sum is negative.

A numerical example makes the penalty concrete. A two-parameter model reaches a best-fit log-likelihood of -52.0; a twenty-parameter model reaches -45.0, better by 7.0. If every parameter of either model is pinned by the data to about 0.3 of its prior range, each contributes $\ln 0.3 = -1.20$. The evidence of the simple model is $-52.0 + 2 \times (-1.20) = -54.4$; that of the flexible model is $-45.0 + 20 \times (-1.20) = -69.1$. The simple model is favored by 14.7 in log evidence, a factor of $e^{14.7} \approx 2.4 \times 10^6$, despite fitting worse.

Figure 16.3 shows the same effect from the side of the data. The horizontal axis is the space of possible measurement outcomes, collapsed to one schematic coordinate; the vertical axis is the probability density each model assigns to each outcome before seeing it. Both curves enclose the same area, one. The rigid model concentrates its probability on a narrow range of outcomes; the flexible model can produce a wider range and must spread its probability thinner to do so. If the observed outcome falls where the rigid model predicted, at the left marker, the rigid model assigns it about twice the density and wins; only an outcome that the rigid model cannot produce, at the right marker, favors the flexible model.

Principle 16.2. *The evidence for a hypothesis is its fit averaged over everything it leaves free. An explanation that fits only after fine tuning of many quantities is penalized relative to one that fits naturally, so a better residual is not by itself a better diagnosis.*

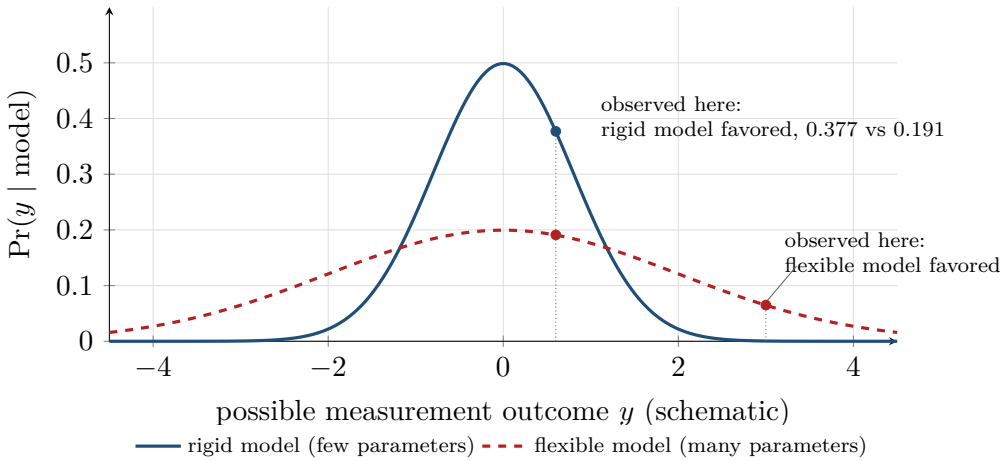


Figure 16.3: Evidence as a prediction of the data, schematic. Each model spreads a total probability of one over the outcomes it can produce: the rigid model as a normal density with standard deviation 0.8, the flexible model with standard deviation 2.0, both centered at zero. At $y = 0.6$ the rigid model's evidence is about twice the flexible model's; at $y = 3$ the order reverses.

16.3.6 Diagnosis as model selection

Each hypothesis can be written as a complete model of the system, with its own free quantities. For the inlet reading of rack 12 three such models are

- $\mathcal{M}_1$: normal airflow and a biased inlet sensor, with the bias b free;
- $\mathcal{M}_2$: reduced airflow and a correct sensor, with the deficit δ free;
- $\mathcal{M}_3$: recirculation and a correct sensor, with the recirculation fraction α free.

Each of them reproduces the inlet reading exactly at its best-fit value: $b = 6\,\mathrm{K}$, $\delta = 0.4$, $\alpha = 0.5$. On the inlet alone, their best fits are identical. Diagnosis is therefore a comparison of models, and the diagnostic posterior is the posterior over models,

$$\Pr(\mathcal{M}_j \mid y) \propto \Pr(y \mid \mathcal{M}_j)\,\Pr(\mathcal{M}_j), \tag{16.8}$$

with each evidence $\Pr(y \mid \mathcal{M}_j)$ computed by equation (16.6). What separates the three is what each predicts for the other sensors, where the best fit of one free quantity can no longer match everything, and how plausible the required value of the free quantity was beforehand: a $6\,\mathrm{K}$ bias is large for a calibrated sensor, a deficit of 0.4 is large for an unobstructed tile, and the evidence weighs both.

16.4 Identifiability, confidence and scale

The posterior is only as sharp as the measurements allow, and the number of hypotheses grows quickly once faults can occur together. This section treats ambiguity and confidence, the combinatorial growth of multiple faults, the factorization that controls it, and the Bayes factors with which evidence overturns a prior.

16.4.1 Ambiguity and confidence

If two hypotheses predict the measurements equally well, $\Pr(y \mid H_1) \approx \Pr(y \mid H_2)$, then the posterior ratio equals the prior ratio,

$$\frac{\Pr(H_1 \mid y)}{\Pr(H_2 \mid y)} = \frac{\Pr(y \mid H_1)}{\Pr(y \mid H_2)} \frac{\Pr(H_1)}{\Pr(H_2)} \approx \frac{\Pr(H_1)}{\Pr(H_2)},$$

and the data have not moved the balance between them. This is not a failure of Bayes' rule. It is the correct statement that the available measurements do not separate the two explanations, the diagnostic form of the identifiability limit of Chapter 15. The honest output is "both explanations remain plausible," together with the measurement that would separate them.

The confidence of a diagnosis is the separation between the leading hypothesis and the rest, not the size of the leading probability on its own. A posterior of 0.51 against 0.47 (with 0.02 left for the others) is unresolved: the odds between the top two are $0.51/0.47 = 1.09$, and the entropy of the distribution is

$$\mathcal{H} = -\sum_i \Pr(H_i \mid y) \log_2 \Pr(H_i \mid y) = 1.12 \text{ bits},$$

more than the one bit of a fair coin. A posterior of 0.97, with 0.02 and 0.01 for the others, is a strong diagnosis: odds of $0.97/0.02 = 48.5$ against the runner-up and an entropy of 0.22 bits. The posterior of Table 16.2, 0.52 against 0.31, lies between the two: recirculation leads, but reduced airflow is far from excluded.

16.4.2 Multiple faults

Faults do not always occur one at a time. A filter can clog while a sensor drifts; a pump can degrade while the load rises. The object of diagnosis is then a joint posterior over combinations, $\Pr(H_A, H_B \mid y)$, and with n faults each either present or absent there are 2^n combinations. Five faults give 32; twenty give

$$2^{20} = 1\,048\,576$$

combinations, each with its own likelihood; fifty give about 1.1×10^{15}. Enumerating them is infeasible beyond a few dozen candidate faults, and the uncertainty itself is combinatorial: probability is spread over many combinations that share most of their members.

What makes the problem tractable is structure. Most faults affect only a few measurements, and most measurements are affected by only a few faults. Inference that exploits this, through factorization, conditional independence, sparse dependency structure and graphical models, never enumerates the full set of combinations.

16.4.3 Factorization and the fault graph

A fault in a UPS affects electrical measurements and has no effect on the sensors of a cooling tower; a fault in a cooling tower affects the condenser water and the chiller and has no effect on UPS voltages. Writing $\mathrm{pa}(j)$ for the set of faults and states that measurement y_j depends on directly, its parents, the likelihood factorizes as

$$\Pr(y \mid H) = \prod_j \Pr(y_j \mid \mathrm{pa}(j)), \tag{16.9}$$

and each factor involves only a few variables. If twenty faults fall into four independent groups of five, the posterior can be computed group by group, $4 \times 2^5 = 128$ combinations instead of 1 048 576. Locality is what makes inference on a large facility possible.

Figure 16.4 draws the structure as a graph. The top row holds the fault hypotheses, the middle row the physical states they act on, and the bottom row the measurements, shaded

because they are observed. An arrow means "directly influences." In the air cluster, a clogged filter and a degraded fan both reduce the airflow $\dot{m}_{\text{air}}$; the airflow sets the rack outlet temperature and the pressure difference; the outlet temperature is read by a sensor whose own bias is a separate fault. The electrical and plant clusters on the right share no arrow with the air cluster.

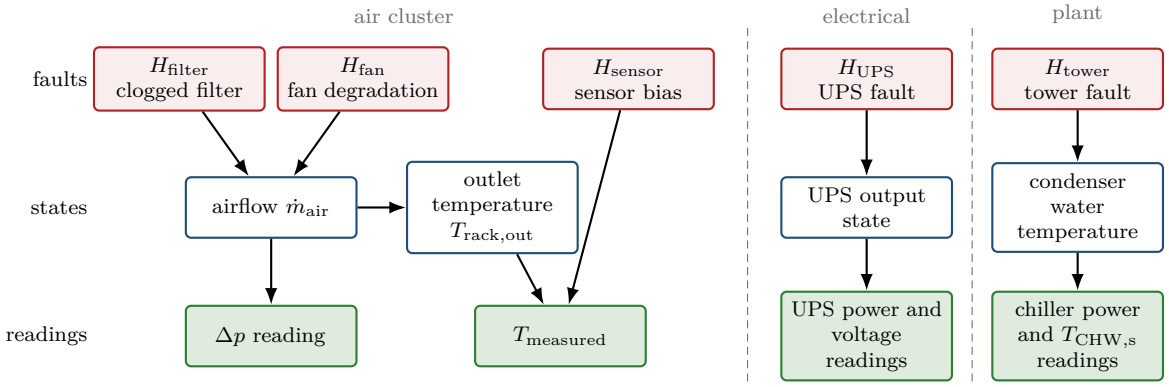


Figure 16.4: A fault graph. Faults (top) act on physical states (middle), which set the readings (bottom, shaded). The likelihood factorizes along the arrows, equation (16.9); clusters with no arrow between them are diagnosed separately, and a reading is evidence only for the faults with a directed path to it.

The graph says which evidence is relevant to which faults. The pressure reading bears on the filter and the fan through the airflow, and on nothing else; the temperature reading bears on the filter, the fan and the sensor. A low airflow therefore cannot separate a clogged filter from a degraded fan, since both act through the same state, and a separating measurement must have a path from one fault that the other lacks, such as the fan's electrical current. Graphs of this kind are Bayesian networks; inference on them uses the factorization of equation (16.9) to pass local messages instead of enumerating combinations.

16.4.4 Bayes factors and evidence against the prior

For two hypotheses, Bayes' rule is simplest in odds form:

$$\frac{\Pr(H_1 \mid y)}{\Pr(H_2 \mid y)} = B_{12}\,\frac{\Pr(H_1)}{\Pr(H_2)}, \qquad B_{12} = \frac{\Pr(y \mid H_1)}{\Pr(y \mid H_2)}, \tag{16.10}$$

posterior odds equal the Bayes factor times the prior odds. If a low airflow reading is ten times more likely under fan degradation than under high IT load, the Bayes factor is $B_{12} = 10$, and from equal prior odds the posterior probability of fan degradation within the pair is $10/(1+10) = 0.91$. Bayes factors of independent readings multiply, by equation (16.3), so several moderate pieces of evidence add up to a decisive one.

Evidence can overturn a prior. Let fan degradation have a prior of $\Pr(H_{\text{fan}}) = 0.01$ and sensor bias a prior of $\Pr(H_{\text{sensor}}) = 0.20$: prior odds of 0.05, twenty to one in favor of the sensor. Three independent readings point to the fan: the airflow estimate, with an illustrative Bayes factor of 10; the pressure, 8; and the fan's electrical current, 12. Their product is $10 \times 8 \times 12 = 960$, the posterior odds are $960 \times 0.05 = 48$, and the posterior probability of fan degradation within the pair is $48/49 = 0.98$. The likelihood has overcome a prior twenty times smaller.

Figure 16.5 plots equation (16.10) as a probability. The horizontal axis is the Bayes factor on a logarithmic scale; the vertical axis is the posterior probability of the hypothesis in the numerator, within the pair. The solid curve starts from equal prior odds, the dashed curve from

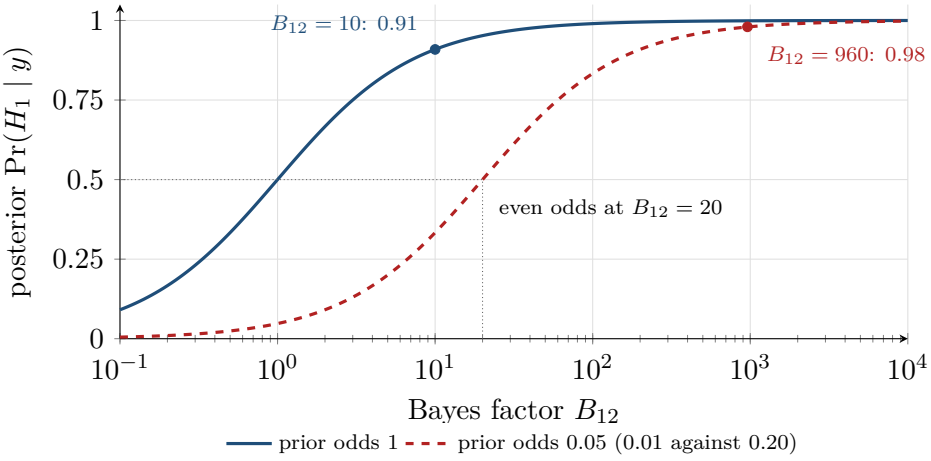


Figure 16.5: Posterior probability within a pair of hypotheses against the Bayes factor, $\Pr(H_1 \mid y) = B_{12}O/(1 + B_{12}O)$ with prior odds $O = 1$ (solid) and $O = 0.05$ (dashed). Markers: a single reading with $B_{12} = 10$ from equal odds; three readings with a combined $B_{12} = 960$ that overturn a twenty-to-one prior.

prior odds of 0.05. The markers are the two examples just worked. The dashed curve shows the cost of a low prior: it needs a Bayes factor of 20 to reach even odds and about 380 to reach 0.95.

> **Principle 16.3.** *Posterior odds are prior odds times the Bayes factor, and the Bayes factors of independent readings multiply. A rare fault becomes the leading explanation when several readings, each modestly in its favor, agree.*

16.5 Time, normal operation and model inadequacy

A diagnosis is made from a record, not from one snapshot, and two hypotheses that every diagnosis should carry are easy to forget: that nothing is wrong, and that the model itself is wrong. This section adds time to the posterior and then adds those two hypotheses.

16.5.1 Persistence of faults

Faults have characteristic durations. A blocked filter does not appear for thirty seconds and vanish; a communication glitch might. Temporal diagnosis encodes this in a transition probability $\Pr(H_t \mid H_{t-1})$, the probability of being in hypothesis H_t at step t given the hypothesis at step $t - 1$. A state that persists from one step to the next with probability q lasts on average $1/(1 - q)$ steps. An illustrative one-minute model for a rack residual uses three states: normal operation, which moves to a transient glitch with probability 0.01 per minute (about one glitch in a hundred minutes) and to a persistent fault with probability 0.002 per minute (about one in eight hours); a transient glitch, which returns to normal with probability 0.5 per minute and so lasts two minutes on average; and a persistent fault, which once present stays.

The distinction between the two kinds of fault is physical. Persistent faults include fouling of a coil, calibration drift of a sensor, a blocked air path and mechanical degradation of a fan or pump: each develops over hours to months and does not reverse on its own. Transient faults include packet loss in the telemetry network, a short power disturbance, an oscillation of a control loop and a temporary spike of workload: each lasts seconds to minutes and clears without intervention. A single abnormal reading fits a transient better than a persistent fault; a sustained shift fits a persistent fault better, and the transition model is what lets the posterior tell the two apart.

16.5.2 Filtering and smoothing

With transitions, the posterior is updated recursively. The filtered posterior uses the measurements up to the present,

$$\Pr(H_t \mid y_{1:t}) \propto \Pr(y_t \mid H_t) \sum_{H_{t-1}} \Pr(H_t \mid H_{t-1}) \Pr(H_{t-1} \mid y_{1:t-1}), \tag{16.11}$$

where $y_{1:t}$ denotes all measurements from step 1 to step t: the previous posterior is carried forward by the transition model and then weighted by the likelihood of the new reading. It is what live diagnosis uses. For a review after an event, the smoothed posterior $\Pr(H_t \mid y_{1:T})$, with $T > t$, also uses the measurements taken after step t; it is computed by a second, backward pass over the record and is the right tool for forensic reconstruction: when did the fault begin, and was an early abnormal reading its first sign or a coincidence? Chapter 14 develops the same recursion for continuous states.

Figure 16.6 applies the three-state model of Section 16.5.1 to a simulated two-hour record. The upper panel is the residual: the measured inlet temperature minus the model's prediction, in kelvin, one reading per minute, with a standard deviation of 0.3 K in normal operation. Two readings at minutes 25 and 26 jump to about 2 K, and from minute 70 the residual is shifted by 0.6 K. The lower panel shows three probabilities against the same time axis: the filtered probability of a persistent fault (solid), the smoothed probability of a persistent fault (dashed) and the filtered probability of a transient glitch (dotted). The glitch puts nearly all the probability on the transient state and leaves the probability of a persistent fault below 0.002: two large readings are what a glitch produces, and the transition model knows that a persistent fault would not stop after two minutes. The persistent shift is smaller per reading but sustained; the filtered probability passes 0.5 at minute 72, two readings after the onset, and the smoothed probability already reaches 0.93 at minute 70, because the later readings show that the shift at minute 70 was the beginning of something that lasted.

16.5.3 Anomaly detection and the normal hypothesis

An anomaly detector computes a score $s(y)$ that grows as the measurements move away from what is usual, for example a distance from the center of historical data scaled by its spread. A high score says "this is unusual." That is useful for detection, but it is not a causal explanation: it does not say which component is responsible, and it does not say whether the unusual reading is abnormal given the present operating conditions. An unusual reading on an unusual day can be entirely normal.

Probabilistic diagnosis addresses the second gap by including a hypothesis H_0, normal operation, with a likelihood computed from the same physical model at the present conditions. Its posterior $\Pr(H_0 \mid y)$ answers the question an alarm threshold cannot: is this behavior abnormal, given the load, the weather and the state of the plant? Including H_0 is what reduces false alarms, because a reading that the normal model predicts is not counted as evidence of a fault however far it is from the facility's typical values.

An illustrative case shows the difference. A high-density rack reads $T_{\text{in}} = 29^{\circ}\text{C}$ and trips a fixed alarm threshold. The rack draws $P_{\text{rack}} = 90\,\text{kW}$, and the cooling plant is behaving as expected. At that load and position, the normal model predicts an inlet near 28.5°C, so a reading of 29°C is only modestly more likely under the fault hypotheses taken together than under H_0: an illustrative Bayes factor of 3.35 in favor of a fault. With a prior of $\Pr(H_0) = 0.95$, the prior odds for normal operation are 19, the posterior odds are $19/3.35 = 5.67$, and

$$\Pr(H_0 \mid y) = \frac{5.67}{1 + 5.67} = 0.85.$$

The reading is unusual in absolute terms and not anomalous in context.

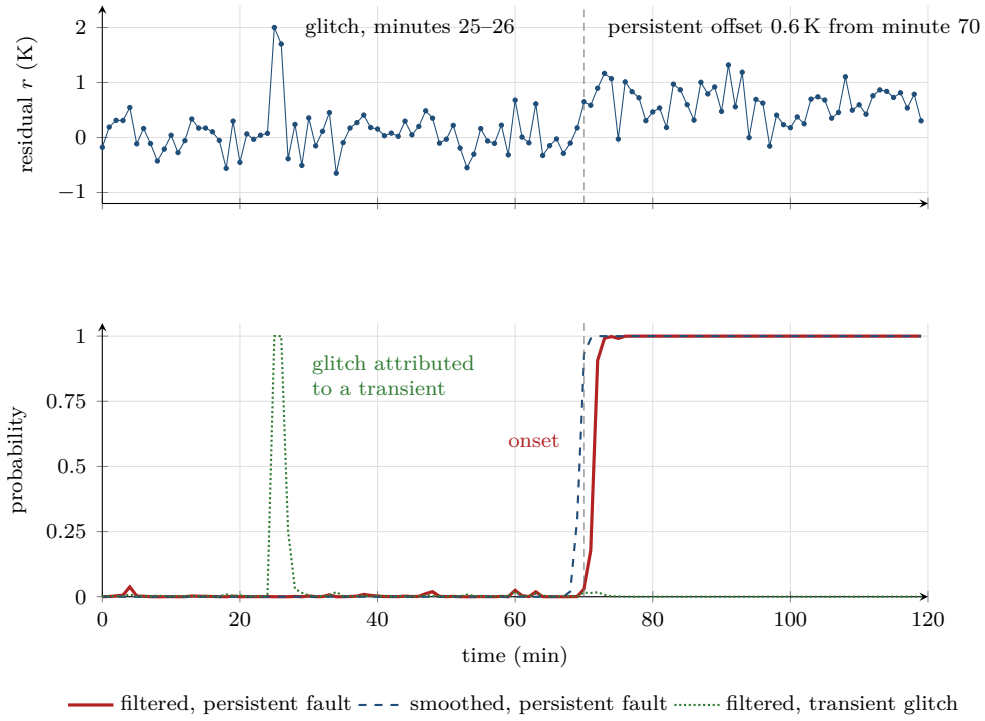


Figure 16.6: Filtering and smoothing on a simulated residual record. Model: normal $r \sim \mathcal{N}(0, 0.3^2)$, transient $r \sim \mathcal{N}(0, 2.0^2)$, persistent $r \sim \mathcal{N}(0.6, 0.3^2)$; per-minute transitions normal to transient 0.01, normal to persistent 0.002, transient to normal 0.5, persistent absorbing. The glitch does not raise the probability of a persistent fault; the sustained offset does, and the smoothed estimate dates its onset two minutes earlier than the filter. Record generated by `scripts/ch16_data.py` with seed 16.

16.5.4 The model-inadequacy hypothesis

Bayes' rule normalizes: the posteriors always sum to one over the hypotheses listed. If every listed hypothesis fits the data badly, normalization still produces a confident-looking ranking. The remedy is a hypothesis H_M, "the model is inadequate," which stands for everything not listed: a missing mechanism, an error in the assumed topology (a rack on a different power path or air path than recorded), an operating mode the model does not represent. H_M makes no sharp prediction; it spreads its probability over a wide range of measurement outcomes, so its likelihood is small but nearly constant. When every candidate fits poorly, $\Pr(y \mid H_i) \ll 1$ for all i, the listed hypotheses fall below that constant and H_M takes the posterior. The honest output is then "none of the current hypotheses adequately explain the data."

Table 16.3 shows the effect with illustrative numbers. Three hypotheses with equal priors have likelihoods in the ratio $5.5 : 3.0 : 1.5$, so without H_M their posteriors are 0.55, 0.30 and 0.15 whatever the absolute size of the likelihoods. In case A the likelihoods are of the size an adequate model produces; in case B they are a million times smaller. H_M receives a prior of 0.02 and a likelihood of 10^{-6} in the same units. The rows give the likelihoods and the posteriors with and without H_M.

Table 16.3: Posteriors with and without a model-inadequacy hypothesis H_M, illustrative. Without H_M the normalized posterior is the same in both cases; with H_M (prior 0.02, likelihood 10^{-6}; the other priors 0.98/3 each) case B is flagged as inadequate.

	H_1	H_2	H_3	H_M
Case A likelihood	5.5×10^{-3}	3.0×10^{-3}	1.5×10^{-3}	10^{-6}
posterior without H_M	0.55	0.30	0.15	–
posterior with H_M	0.55	0.30	0.15	6×10^{-6}
Case B likelihood	5.5×10^{-9}	3.0×10^{-9}	1.5×10^{-9}	10^{-6}
posterior without H_M	0.55	0.30	0.15	–
posterior with H_M	0.077	0.042	0.021	0.860

The adequacy test. The same check can be made directly, before any ranking. If the best hypothesis leaves standardized residuals whose sum of squares follows a chi-square distribution with as many degrees of freedom as independent readings, its value can be compared with that distribution. With five readings, a sum of squares of 4.2 is ordinary (a probability of 0.52 of a value at least that large); a sum of 38 lies far beyond the 95 % point of 11.07, with a probability of about 4×10^{-7}, and says that the best available explanation does not explain the data. Without such a test, a posterior of $(0.55, 0.30, 0.15)$ looks equally confident in both cases of Table 16.3. An adequacy test precedes the posterior ranking, and a ranking is reported only for a model that passes it.

> **Principle 16.4.** *A posterior ranks only the hypotheses that were listed. Before it is reported, the best of them must be tested for whether it explains the data at all; if none does, the correct diagnosis is that the model is inadequate.*

16.6 From posterior to action

A posterior is useful when it changes what is done. This section weighs hypotheses by their consequences, chooses actions by expected loss, values a measurement by the uncertainty it

removes, designs safe experiments, and assembles the result into an operator's diagnostic output and one joint formulation.

16.6.1 Risk ranking

The most probable explanation is not necessarily the most important one. Each hypothesis has a consequence or loss L_i if it is true and left unattended, and its contribution to risk is the product

$$\mathcal{R}_i = \Pr(H_i \mid y)\, L_i. \tag{16.12}$$

Table 16.4 applies equation (16.12) to an illustrative posterior for a cooling symptom, with losses in relative units: a sensor bias costs 1 (a wrong reading, a recalibration), a partial loss of airflow costs 5 (local overheating, throttled servers), and a degrading chiller costs 20 (a plant-wide loss of cooling margin and possibly an outage). The columns are the posterior, the loss and their product.

Table 16.4: Risk ranking, illustrative. The most probable hypothesis carries the smallest risk; the least probable carries the largest.

Hypothesis	$\Pr(H_i \mid y)$	loss L_i	risk $\mathcal{R}_i$
Sensor bias	0.55	1	0.55
Partial airflow loss	0.30	5	1.50
Chiller degradation	0.15	20	3.00
Total	1.00		5.05

Sensor bias is the most likely explanation; chiller degradation carries the highest risk, 3.00 of a total of 5.05, and the chiller is the first thing to check. A ranking by probability alone would send the technician to the sensor and leave the largest exposure unexamined.

16.6.2 Expected-loss decisions

Decision theory makes the step from posterior to action explicit. Let $L(a, H)$ be the loss of taking action a when hypothesis H is true. The best action is the one with the smallest expected loss under the posterior,

$$a^{*} = \arg\min_{a} \; \mathbb{E}\big[L(a, H) \mid y\big] = \arg\min_{a} \; \sum_{i} L(a, H_i)\, \Pr(H_i \mid y). \tag{16.13}$$

This is not the same as acting on the most likely fault. The most likely fault ignores both how wrong the action is when another hypothesis is true and what that error costs.

An illustrative example has two hypotheses: H_1, sensor bias, with posterior 0.6, and H_2, a real loss of cooling, with posterior 0.4. Three actions are available: do nothing; act now, by starting a standby cooling unit and moving workload away from the affected racks; and measure first, with an independent handheld temperature measurement that takes ten minutes, then act if the loss is confirmed. Table 16.5 lists illustrative losses in thousands of dollars. Doing nothing costs nothing under H_1 and 100 under H_2, an overheating event. Acting now costs 8 under H_1, an unnecessary intervention, and 10 under H_2, the intervention plus a small residual effect. Measuring first costs 1 under H_1 and $1 + 8 + 2 + 2 = 13$ under H_2: the measurement, the intervention, the residual effect and the extra effect of ten minutes' delay.

Measuring first has the smallest expected loss, 5.8 against 8.8 for acting now and 40.0 for doing nothing. Doing nothing, the action suggested by the most likely hypothesis, is the worst of the three: cheap under H_1 and disastrous under H_2. Verification is cheap under either hypothesis and removes most of the uncertainty, which is why it wins.

Table 16.5: Losses $L(a, H)$ in thousands of dollars and expected losses under the posterior $\Pr(H_1 \mid y) = 0.6$, $\Pr(H_2 \mid y) = 0.4$, illustrative.

Action a	H_1 sensor bias	H_2 cooling loss	expected loss
Do nothing	0	100	$0.4 \times 100 = 40.0$
Act now	8	10	$0.6 \times 8 + 0.4 \times 10 = 8.8$
Measure first	1	13	$0.6 \times 1 + 0.4 \times 13 = 5.8$

Figure 16.7 shows how the answer depends on the posterior. The horizontal axis is the probability of a real cooling loss, $p = \Pr(H_2 \mid y)$; the vertical axis is the expected loss of each action, a straight line in p: $100p$ for doing nothing, $8 + 2p$ for acting now and $1 + 12p$ for measuring first. The lowest line at each p is the best action. Doing nothing is best only for $p < 1/88 \approx 0.011$; measuring first is best from there up to $p = 0.7$, where $1 + 12p = 8 + 2p$; above 0.7 the case for a cooling loss is strong enough that the delay of measuring costs more than it saves, and acting at once is best. The posterior of the example, 0.4, lies well inside the measure-first range.

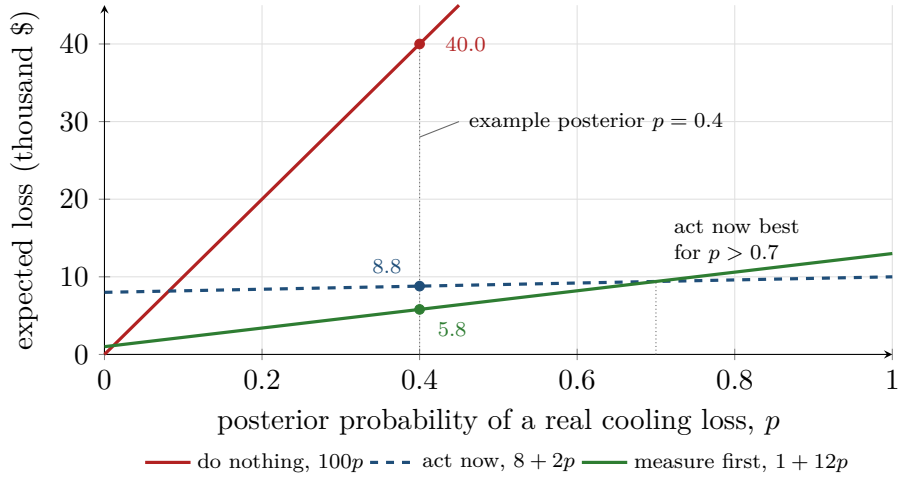


Figure 16.7: Expected loss of three actions against the posterior probability of a real cooling loss, from the illustrative losses of Table 16.5. The lowest line is the best action: do nothing below $p = 0.011$, measure first between 0.011 and 0.7, act now above 0.7. The do-nothing line continues to 100 at $p = 1$ and is cut at the top of the axis.

16.6.3 Value of information of a measurement

When the best action is to measure, the question becomes which measurement. In the rack 12 case the leading hypotheses were recirculation and reduced airflow, and three candidate measurements are available: m_1, a direct measurement of the rack airflow with a portable flow hood; m_2, a local hot-aisle temperature directly behind the rack; and m_3, the pressure drop across the rack's air filter. One way to value a measurement m is by the expected reduction in the entropy of the posterior,

$$\mathrm{VOI}(m) = \mathcal{H}(H \mid y) - \mathbb{E}_m\big[\mathcal{H}(H \mid y, m)\big], \tag{16.14}$$

where $\mathcal{H}(H \mid y)$ is the entropy of the present posterior and the expectation runs over the outcomes the measurement could produce, each weighted by its predicted probability. This entropy reduction is also called the information gain IG of the measurement.

Table 16.6 evaluates equation (16.14) for the posterior of Table 16.2, rounded to two decimals: 0.04, 0.31, 0.52, 0.01 and 0.12 for H_1 to H_5, with an entropy of 1.63 bits. Illustrative likelihoods: m_1 reads low with probability 0.9 under reduced airflow and 0.1 under every other hypothesis; m_2 reads high with probability 0.8 under H_1, H_3 and H_4, 0.7 under H_2 and 0.2 under sensor bias; m_3 reads high with probability 0.5 under reduced airflow (when the filter is the cause) and 0.1 otherwise. Each row gives the probability of the outcome that favors reduced airflow, the posterior entropy after each outcome, their weighted mean and the resulting value of information.

Table 16.6: Entropy-based value of information of three candidate measurements for rack 12, illustrative likelihoods. Present entropy 1.63 bits.

Measurement	Pr(1st outcome)	$\mathcal{H}$ after 1st	$\mathcal{H}$ after 2nd	expected $\mathcal{H}$	VOI (bits)
m_1 rack airflow (low / normal)	0.348	0.93	1.30	1.17	0.46
m_2 local hot aisle (high / normal)	0.697	1.41	1.76	1.52	0.11
m_3 filter Δp (high / normal)	0.224	1.22	1.58	1.50	0.13

The rack airflow is worth about four times either alternative. It bears directly on the one quantity on which the two leading hypotheses disagree: a low reading raises reduced airflow to 0.80, a normal reading raises recirculation to 0.72. The local hot-aisle temperature mostly repeats what the aisle reading already said, since both leading hypotheses predict a warm hot aisle, and the filter pressure drop is informative only in the sub-case where the filter is the cause. Chapter 19 develops the choice of the next measurement in full, including values of information measured in expected loss rather than in bits.

16.6.4 Active diagnosis and its constraints

Measurements need not be passive. An operator can perturb the system deliberately, by raising a fan speed slightly, changing a valve position, altering the IT load or switching to a redundant cooling path, and the hypotheses predict different responses. This is active diagnosis.

An illustrative test separates H_1, fan degradation, from H_2, an airflow obstruction. Both produce the same reduced airflow in steady operation, so the temperatures cannot separate them. The test raises the fan speed command by 10 %. A simple fan model, a pressure rise of $40(N/N_0)^2 - 5\dot{V}^2$ in pascals against a system resistance $\Delta p = K\dot{V}^2$ (Chapter 6), gives the responses of Table 16.7. A healthy path has $K = 5\,\mathrm{Pa\,s^2/m^6}$. Under H_1 the fan already runs below its commanded speed and, at its limit, follows the 10 % command by only 2 %. Under H_2 the fan is healthy and the path resistance has risen to $K = 12.5$. The rows give the airflow and the pressure across the fan before and after the command change, and the resistance $\Delta p/\dot{V}^2$ inferred from them.

Table 16.7: Responses to a 10 % increase in fan speed command, illustrative fan and system curves. Airflow in $\mathrm{m^3/s}$, pressure in Pa.

Case	$\dot{V}$ before	$\dot{V}$ after	Δp before	Δp after	$\Delta p/\dot{V}^2$
Healthy path	2.00	2.20	20.0	24.2	5.0
H_1 fan degradation	1.51	1.54	11.4	11.9	5.0
H_2 obstruction	1.51	1.66	28.6	34.6	12.5

The two signatures are distinct. Under fan degradation the airflow barely changes, by

$+0.03\,\mathrm{m}^3/\mathrm{s}$, and neither does the pressure, because the fan cannot deliver what is commanded. Under obstruction the pressure rises strongly, $+6.0\,\mathrm{Pa}$, more than the $+4.2\,\mathrm{Pa}$ of a healthy path, while the flow stays constrained, $+0.15\,\mathrm{m}^3/\mathrm{s}$ against $+0.20$; the inferred resistance is two and a half times the healthy value.

Experiments must be safe. A perturbation is permitted only if it keeps the system within its limits under every hypothesis still plausible: the rack inlet temperature below its limit, $T_{\mathrm{in}} \le T_{\max}$; the facility power within its limit, $P_{\mathrm{facility}} \le P^{\max}$; and the service commitments of the IT load satisfied. Raising a fan speed by $10\,\%$ raises fan power by about $1.1^3 - 1 = 33\,\%$ by the fan laws, which must fit within the power limit; closing a valve or shifting load can raise temperatures under the very hypothesis being tested. Choosing a test is therefore a constrained optimization: among perturbations u, choose the one with the largest expected information gain subject to $\Pr(T_{\mathrm{in}} \le T_{\max}$ throughout the test $\mid y) \ge 1 - \epsilon$ and $P_{\mathrm{facility}} \le P^{\max}$, with ϵ a small tolerated risk. Chapter 20 treats optimization under such constraints.

16.6.5 The diagnostic output and the joint formulation

A practical diagnostic output collects the result in a form an operator can act on. Table 16.8 shows one for rack 12: the observed condition, the leading hypotheses with their probabilities, the evidence that supports them, the present risk stated as a probability, and the recommended next measurement. The risk entry, a probability of exceeding 35°C within fifteen minutes, is computed from the posterior over states and is the subject of Chapter 17.

Table 16.8: A diagnostic output for rack 12, illustrative.

Observed condition	Rack 12 inlet temperature elevated (29°C against a normal 23°C).
Leading hypotheses	$\Pr(H_{\mathrm{recirc}} \mid y) = 0.52$; $\Pr(H_{\mathrm{airflow}} \mid y) = 0.31$.
Why	Supply temperature normal; local hot-aisle temperature elevated; cold-aisle differential pressure low; rack power normal.
Current risk	$\Pr(T_{\mathrm{in}} > 35°\mathrm{C}$ within 15 min $\mid y) = 0.08$.
Recommended next measurement	Rack airflow (value of information 0.46 bits, about four times the alternatives).

Such an output does more than a monitoring dashboard. A dashboard shows temperatures, alarms and trends. An anomaly detector adds that rack 12 is unusual. A probabilistic physical diagnosis adds what is probably happening, why, how uncertain the conclusion is, what risk it implies, and what to measure or do next.

All of the chapter's quantities come from one joint posterior over hypotheses, states and parameters,

$$p(H, x, \theta \mid y) \propto p(y \mid H, x, \theta)\, p(x, \theta \mid H)\, \Pr(H). \tag{16.15}$$

From it follow the diagnosis $\Pr(H \mid y)$, by integrating out x and θ; the state estimate $p(x \mid y)$ and the parameter estimate $p(\theta \mid y)$, by summing over H; the risk $\Pr(g(x, \theta, H) > 0 \mid y)$, where $g > 0$ marks a violation such as an inlet above its limit; and the decision $a^* = \arg\min_a \mathbb{E}[L(a, H, x) \mid y]$.

Figure 16.8 draws the structure. The three boxes on the left are the inputs: the measurements, the physical model that gives the likelihood, and the priors over hypotheses, states and parameters. The central box is the joint posterior of equation (16.15). The five boxes on the right are what is read from it. The arrow along the bottom closes the loop: the decision is often a measurement or a test, whose result returns as new evidence. The chain from left to right, evidence, competing explanations, posterior probabilities, consequences, action or next measurement, is the procedure of the whole chapter.

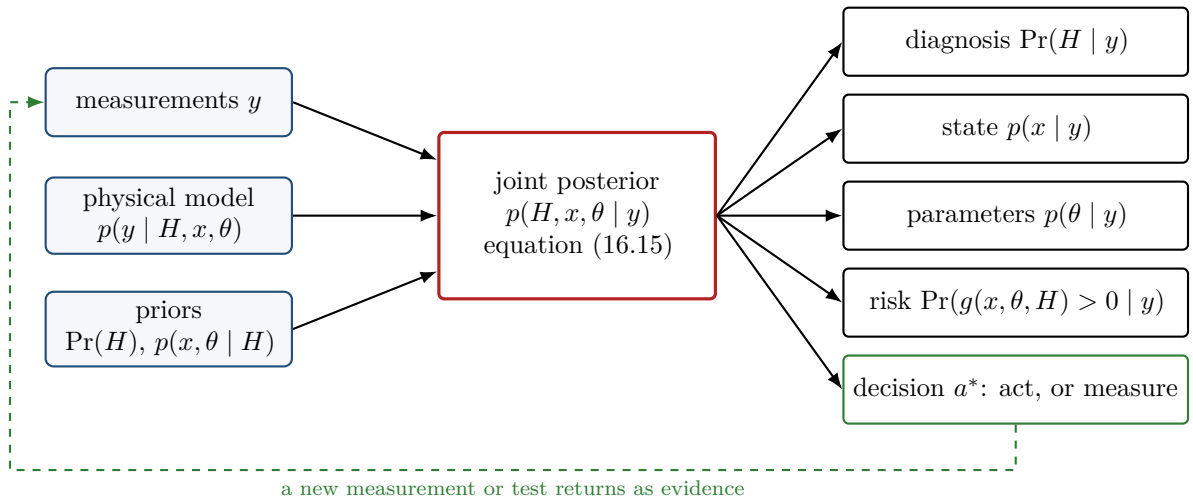


Figure 16.8: The joint formulation. Measurements, a physical model and priors define the joint posterior over hypotheses, states and parameters; diagnosis, state and parameter estimates, risk and the decision are all read from it. When the decision is to measure or to test, the result re-enters as evidence (dashed).

16.7 What to retain

Definition 16.1 (Probabilistic diagnosis)**.** Probabilistic diagnosis maintains a physically consistent probability distribution over competing explanations of the measurements, $\Pr(H_i \mid y) \propto \Pr(y \mid H_i)\,\Pr(H_i)$, with likelihoods computed from a physical model and averaged over unmeasured states and parameters, a normal-operation and a model-inadequacy hypothesis among the candidates, and uses it to rank risks and to choose the action or measurement with the smallest expected loss.

Table 16.9 lists the chapter's objects and the subsection that defines each.

16.8 Exercises

Exercise 16.1. A CRAH unit reports a supply temperature 3 K above its setpoint. Three hypotheses are considered: H_1, a chilled-water valve stuck partly closed (prior 0.15); H_2, a fouled coil (prior 0.25); H_3, a biased supply sensor (prior 0.60). The likelihoods of the supply reading are 0.7, 0.5 and 0.4. (a) Compute the posterior of each hypothesis. (b) A second reading is added: the chilled-water flow through the unit, which is low with probability 0.8 under H_1, 0.2 under H_2 and 0.2 under H_3, and is observed low. Assuming conditional independence, compute the new posterior. (c) Compute the entropy before and after the second reading, and say whether the diagnosis is resolved.

Exercise 16.2. A pump degradation has prior 0.02 and a flow-meter drift has prior 0.25. (a) Compute the prior odds of degradation against drift. (b) Three independent readings give Bayes factors of 5, 6 and 8 in favor of degradation. Compute the posterior odds and the posterior probability of degradation within the pair. (c) How large a combined Bayes factor is needed to reach a posterior of 0.95? (d) If each additional reading contributes a Bayes factor of 4, how many readings beyond the first three are needed?

Exercise 16.3. For the two hypotheses of Table 16.5, replace the losses by: do nothing, 0 and 150; act now, 12 and 14; measure first, 2 and 18. (a) Write the expected loss of each action as a function of $p = \Pr(H_2 \mid y)$. (b) Find the ranges of p in which each action is best. (c) Which action is best at $p = 0.4$, and by how much does it beat acting on the most likely hypothesis?

Table 16.9: The objects of probabilistic diagnosis and where each is defined.

Object	Expression	Role	Defined in
Bayes' rule	equation (16.1)	posterior from prior and likelihood	16.2.1
Likelihood product	equation (16.3)	conditionally independent readings	16.2.4
Physical likelihoods	equation (16.4), $\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}} c_{p,\text{air}} \Delta T$	forward predictions per hypothesis	16.3.1
Prior	$\Pr(H_i)$	plausibility before the data	16.3.3
Marginal likelihood	equations (16.5), (16.6)	averages over states and parameters	16.3.4
Occam penalty	equation (16.7)	penalizes tuned flexibility	16.3.5
Model posterior	$\Pr(\mathcal{M}_j \mid y)$	diagnosis as model selection	16.3.6
Entropy	$\mathcal{H}$ in bits	confidence as separation	16.4.1
Factorization	equation (16.9)	tractable multi-fault inference	16.4.3
Bayes factor	B_{12}, equation (16.10)	evidence against a prior	16.4.4
Transition model	$\Pr(H_t \mid H_{t-1})$	persistent versus transient	16.5.1
Filter, smoother	equation (16.11), $\Pr(H_t \mid y_{1:T})$	live and forensic diagnosis	16.5.2
Normal hypothesis	H_0	abnormal given conditions	16.5.3
Inadequacy hypothesis	H_M, chi-square test	none of the hypotheses fits	16.5.4
Risk	$\mathcal{R}_i = \Pr(H_i \mid y) L_i$	probability times consequence	16.6.1
Expected-loss decision	equation (16.13)	best action under uncertainty	16.6.2
Value of information	equation (16.14)	which measurement next	16.6.3
Active test	Table 16.7	safe perturbation that separates	16.6.4
Joint posterior	equation (16.15)	source of every output	16.6.5

Exercise 16.4. Two hypotheses remain, recirculation with posterior 0.6 and reduced airflow with 0.4. (a) Compute the entropy in bits. (b) A rack airflow measurement reads low with probability 0.9 under reduced airflow and 0.1 under recirculation. Compute the probability of each outcome, the posterior after each, the expected posterior entropy and the value of information of equation (16.14). (c) A pressure measurement reads low with probability 0.6 under recirculation and 0.4 under reduced airflow. Compute its value of information and say which measurement to take. (d) With the fan model of Table 16.7, what airflow and fan pressure would a path with $K = 20$ show at the original speed?

16.9 Connection to the TERRA framework

Probabilistic diagnosis is the inference half of the framework applied to discrete alternatives. Each hypothesis is a variant of the same graph, and every likelihood of the chapter is a forward solution of that graph.

- **A hypothesis is a modified graph.** H_3 adds an air edge from the hot-aisle node to the rack 12 inlet node with recirculation fraction α, and equation (16.4) is the mass and energy balance of that inlet node; H_2 lowers the flow on the supply edge into the cold-aisle node; H_1 raises the power flow on the rack's electrical edge, which the heat balance turns into $\dot{Q}_{\text{rack}}$; H_5 adds a bias b to one measurement edge. The signatures of Table 16.1 are the forward solutions of these five graphs.
- **The likelihood is marginal over inferred quantities.** Rack airflow, the recirculation fraction, coil conductance and sensor bias are inferred, not measured, so each hypothesis's likelihood is equation (16.6): posterior inference of x and θ under that graph, with the evidence approximated around the posterior mode as in equation (16.7).

- **Graph locality is the factorization.** A measurement is attached to a node or an edge, and a fault changes a component on one of the three coupled networks, electrical, air and water mass, and thermal. The fault graph of Figure 16.4 is read off that structure, and equation (16.9) is the statement that a UPS fault and a tower fault touch disjoint measurement sets.

- H_0 **and** H_M **are residual tests on balances.** Normal operation is the unmodified graph at the present load and weather; the balances that should close (facility power, air side against water side, evaporator against condenser) give residuals whose chi-square adequacy is checked before any ranking, and a failed test points to a missing component or a topology error, such as a rack on a different power path than recorded.

- **Risk and decision are read from the posterior.** The risk entry of Table 16.8 is the probability that the available cooling at the rack 12 node falls below the required cooling, $\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}} \mid y)$, computed by forward simulation in time from the joint posterior of equation (16.15); the next measurement is the one with the largest expected reduction in the entropy of that posterior.

In summary, a diagnosis is a ranked posterior over physically defined explanations, computed by Bayes' rule from priors that track the facility's history and likelihoods that come from forward physics averaged over what is not measured; it is sharp only where the measurements separate the hypotheses, it must include normal operation and model inadequacy among its candidates, and it becomes useful when its probabilities are weighed by consequences and turned into the action or measurement with the smallest expected loss. The next chapter turns the same posterior into a statement of capacity and risk: the probability that the cooling available to a rack or a room falls short of what it requires (Chapter 17).

Chapter 17

Capacity and Risk

17.1 Capacity as a system property

The preceding chapters described how a data center delivers power, removes heat, measures itself, fails, and is diagnosed. This chapter asks the planning and operating question that rests on all of them:

how much additional IT load can the facility support safely, under the conditions that hold now and in the hours ahead?

The obvious answer, installed capacity minus current load, is wrong in almost every facility. The load the facility can actually carry is limited by the electrical distribution, by the cooling plant, by the airflow that reaches each rack, by floor space and network ports, and by the reliability margin that must be held back for failures. Each of these limits depends on the operating state, several of them depend on quantities that are inferred rather than measured, and all of them are uncertain to some degree. This section separates installed capacity from usable capacity and defines the margin; the sections that follow make capacity depend on weather and contingencies, on location in the room, on uncertainty, on time, and finally turn it into a planning instrument.

Jargon: nameplate capacity, usable capacity, headroom, capacity margin, bottleneck, contingency, N−1, secure capacity, chance constraint, quantile capacity, stranded capacity

Nameplate capacity is the rating printed on the equipment or stated in the design: the megawatts a transformer, a UPS system or a chiller plant is built to carry. *Usable capacity* is the load the facility as a whole can actually support with every limit respected. *Headroom* is the colloquial name for the unused part of a capacity, and the *capacity margin* $M = C - L$ is its precise form: capacity C minus load L, both in the same units. A *bottleneck* is the subsystem whose limit is reached first and therefore sets the usable capacity. A *contingency* c is a credible failure or outage, such as the loss of one transformer, one UPS module or one air handler; *N−1* means that the system must remain within its limits after any single contingency. *Secure capacity* is the largest load that can be carried while every required contingency remains survivable. A *chance constraint* is a limit stated as a probability, "the probability of a violation must stay below ϵ", rather than as a fixed inequality. A *quantile capacity* C_{95} is the value the uncertain capacity exceeds with probability 0.95. *Stranded capacity*, introduced briefly in Chapter 11, is installed capacity that cannot be used because another subsystem binds first.

Key idea

Usable capacity is not installed capacity. It is the smallest of several subsystem limits, each of which moves with weather, equipment state and contingencies and is known only with uncertainty. The useful statement of capacity is the additional load the system can support under current conditions while the probability that load exceeds capacity, $\Pr(M < 0)$, stays acceptably low.

17.1.1 Nameplate and usable capacity

Consider an illustrative facility with 100 MW of electrical infrastructure: utility transformers, switchgear, UPS systems and power distribution units sized for 100 MW of IT load. Its nameplate electrical capacity is 100 MW. Its other limits are lower. The cooling plant can remove the heat of 85 MW of IT load. Under the N−1 requirement, the electrical system must survive the loss of any single path, and with one path out it can carry only 80 MW. In some zones of the floor the airflow cannot be delivered to the rack positions that remain, so the IT load that can actually be deployed in those zones restricts the whole facility to 75 MW. The usable capacity is the smallest of these,

$$\begin{aligned} C_{\text{usable}} &= \min(C_{\text{electrical}},\ C_{\text{cooling}},\ C_{\text{airflow}},\ C_{\text{reliability}},\ \dots) \\ &= \min(100, 85, 75, 80)\,\text{MW} = 75\,\text{MW}, \end{aligned} \tag{17.1}$$

with every capacity expressed in megawatts of IT load, so that the limits can be compared directly. The list continues with floor space, rack positions, network ports and uplinks, and the generator plant; each supports some maximum IT load, and the smallest wins.

Figure 17.1 draws the four limits of the example as horizontal bars on a common scale of megawatts of IT load. The dark part of each bar is the capacity up to the usable value; the light part is the capacity that exists but cannot be used. The dashed vertical line is the usable capacity, 75 MW, and the red bar is the bottleneck, the local airflow. The figure shows that 25 MW of the nameplate electrical capacity, a quarter of it, is unusable without any equipment being faulty: the facility is built for 100 MW and can host 75 MW.

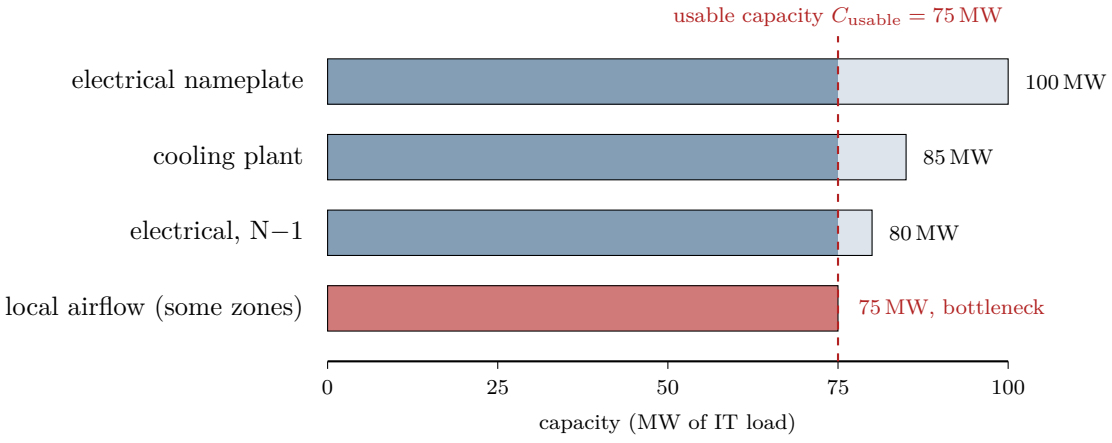


Figure 17.1: Subsystem limits of an illustrative 100 MW facility, all in megawatts of IT load. Usable capacity is the minimum, equation (17.1): 75 MW, set by the local airflow. The light part of each bar is installed capacity that cannot be used.

Usable capacity is therefore a property of the system, not of any one piece of equipment. It cannot be read off a nameplate, and it changes when any limit changes: a chiller taken out for maintenance, an airflow path blocked by new cabling, a UPS module that fails, all lower one

of the terms of equation (17.1), and the minimum moves whenever the term that moves is, or becomes, the smallest.

Principle 17.1. *Usable capacity is the minimum over the subsystem limits, all expressed in the same units of IT load. It is a system-level property, set by the bottleneck, and it moves whenever the binding limit moves.*

17.1.2 The capacity margin

The quantity that matters operationally is not capacity alone but what is left of it after the present load. The capacity margin is

$$M = C - L, \tag{17.2}$$

with C a capacity and L the load placed on it, in the same units. For an illustrative cooling system with $C_{\text{cool}} = 10\,\text{MW}$ carrying a heat load $L_{\text{cool}} = 8\,\text{MW}$, the margin is

$$M = 10\,\text{MW} - 8\,\text{MW} = 2\,\text{MW}.$$

A positive margin means the load can grow by that amount before the limit is reached; a negative margin means the limit is already exceeded. The margin can be written for any subsystem, the cooling plant, one electrical path, one rack, and for the facility as a whole through C_{usable}.

The number 2 MW is incomplete until it states the conditions under which it holds. The 10 MW capacity was computed at some outdoor temperature, with some set of chillers available, with coils in some state of cleanliness and with some airflow distribution. The 8 MW load was measured at one moment. If any of those conditions changes, the margin changes, and the rest of the chapter is about how: through the operating state (Section 17.2), through location in the room (Section 17.3), through uncertainty (Section 17.4) and through time (Section 17.5). In the book's vocabulary, the load is measured, the capacity is calculated from a model of the equipment, and several inputs to that model, the airflow reaching a rack or the conductance of a fouled coil, are inferred; the margin inherits the uncertainty of the least certain of them.

17.2 State-dependent and secure capacity

A capacity is not a constant of the facility. This section shows how cooling capacity depends on weather and equipment state, how electrical capacity depends on the contingency the system must survive, and how the two combine into secure capacity, the largest load that remains feasible after the failures the facility is designed for.

17.2.1 Dependence on operating conditions

The capacity of the cooling system is a function of the conditions in which it operates,

$$C_{\text{cool}} = f\big(T_{\text{amb}},\ \text{RH},\ \text{equipment availability},\ \text{fouling},\ \dot{m},\ \text{control state}\big), \tag{17.3}$$

and the capacity of the electrical system is a function of its own,

$$C_{\text{elec}} = f\big(G,\ \text{transformer loading},\ \text{UPS availability},\ c\big), \tag{17.4}$$

where G is the topology, which source feeds which bus and which bus feeds which rack, and c the contingency state, which components are out of service. Each argument enters through physics established in earlier chapters. Outdoor dry-bulb and wet-bulb temperatures set the condenser temperature and with it a chiller's coefficient of performance and maximum duty

(Chapter 7); fouling lowers the overall conductance UA of a coil and with it the heat the coil can transfer at a given temperature difference; the mass flows $\dot{m}$ of air and water bound the heat a loop can carry through $\dot{Q} = \dot{m}\, c_p\, \Delta T$ (Chapter 5); the control state, the setpoints, the number of units staged on, and whether an economizer is active, decides which of these limits applies. On the electrical side, a transformer's continuous rating depends on its loading history and ambient temperature, and the capacity of a path depends on whether its redundant partner is in service.

Capacity is therefore a state-dependent quantity, in the same sense that the available power of a battery depends on its state of charge and temperature. A capacity quoted without its state is a capacity at an unstated state, usually the design condition, and the design condition is not the condition of today.

17.2.2 Weather and cooling capacity

Take an illustrative cooling plant that can remove 12 MW of heat at an outdoor temperature of 20°C and only 9.5 MW at 38°C with high humidity, because the condensers reject heat into hotter, wetter air and the chillers reach their maximum compressor power at a smaller duty. The IT heat load is 8.5 MW in both cases. The margins are

$$M_{20^\circ\mathrm{C}} = 12 - 8.5 = 3.5\,\mathrm{MW}, \qquad M_{38^\circ\mathrm{C}} = 9.5 - 8.5 = 1.0\,\mathrm{MW}.$$

The margin has fallen by 2.5 MW, more than two thirds of its mild-day value, with no change of equipment, no fault and no change of load. A facility that reports "3.5 MW of cooling headroom" from a measurement taken on a mild morning is reporting a number that the afternoon will not honor.

The left panel of Figure 17.2 draws this example. The horizontal axis is the outdoor temperature and the vertical axis is power in megawatts. The solid line is the cooling capacity, drawn as a straight line through the two stated points, which is an illustrative interpolation; the dashed horizontal line is the IT heat load. The two vertical arrows are the margins at 20 and 38°C. The capacity line and the load line converge as the weather warms; extended, they would meet near 45°C, the temperature at which this plant would run out of margin at this load.

17.2.3 Contingencies and electrical capacity

An illustrative electrical path rated 10 MW carries 7 MW in normal operation, a margin of 3 MW. A second path, which carries 4 MW of its own load, fails, and its load transfers to the first path, as it is designed to. The first path now carries

$$7 + 4 = 11\,\mathrm{MW}, \qquad M_{\mathrm{N}-1} = 10 - 11 = -1\,\mathrm{MW}.$$

The normal-state margin of 3 MW was real, and it was not enough: the path must hold in reserve the load it may receive from a failed partner. The right panel of Figure 17.2 shows the two states. The horizontal axis separates the normal state from the state after the partner path is lost; the vertical axis is the load on the first path. The dashed line is the 10 MW rating. The blue part of each bar is load within the rating, the red part is load above it. A capacity statement of the form "this path has 3 MW unused" is a statement about the normal state only; the operationally relevant figure is the margin in the worst credible contingency, which here is negative.

This is the same reasoning as contingency analysis of a transmission system, where a line that is lightly loaded in the base case may be the one that overloads when a parallel line trips. Chapter 3 described the redundant power paths that make such transfers possible and Chapter 12 develops their reliability; here the point is only that the capacity of a path depends on which contingency is assumed.

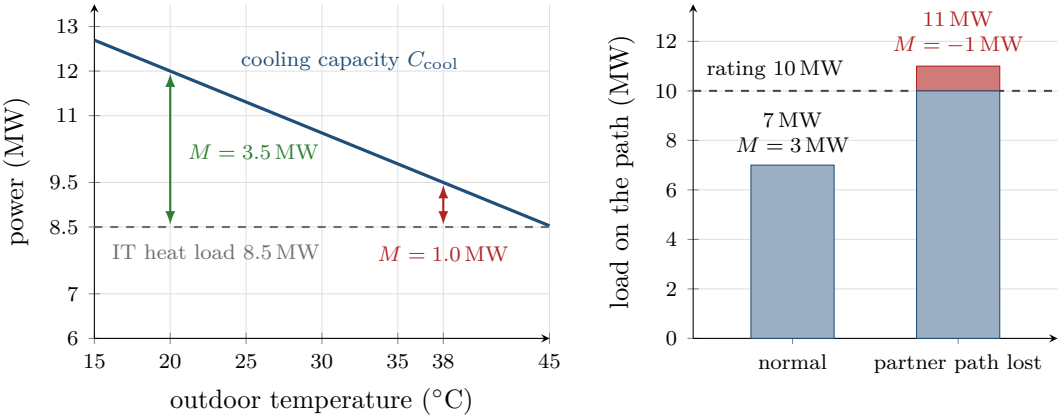


Figure 17.2: State dependence of capacity, illustrative. Left: cooling capacity against outdoor temperature, the straight line $C_{\text{cool}} = 12 - (2.5/18)(T_{\text{amb}} - 20)$ MW through the two stated points, and an IT heat load of 8.5 MW; the margin falls from 3.5 to 1.0 MW between 20 and 38°C. Right: a 10 MW path carrying 7 MW receives 4 MW when its partner path fails; its margin goes from +3 to −1 MW.

17.2.4 Secure capacity

The two examples generalize into a definition. The *secure capacity* of a facility is the largest load it can carry while every limit holds in the normal state and in every contingency of a required set $\mathcal{C}$,

$$C_{\text{secure}} = \max\ L \quad \text{subject to all limits hold for } L \text{ in state } c \text{ for every } c \in \mathcal{C} \cup \{\text{normal}\}. \tag{17.5}$$

The required set is a design or policy choice: all single failures for an N−1 facility, selected double failures for a facility with a stricter requirement, and in either case only the contingencies judged credible. The concept is the same as the secure transfer capacity of a transmission corridor, the flow that can be scheduled while the corridor still survives its worst credible outage.

For an illustrative facility whose nameplate capacity is 100 MW, the binding contingency, say the loss of its largest transformer with that transformer's load transferred to the remaining ones, leaves $C_{\text{secure}} = 82$ MW. The facility is a 100 MW facility on paper and an 82 MW facility in operation; the 18 MW difference is the price of being able to survive the contingency, and it is capacity that must not be sold. (This is a different illustrative facility from the one of Figure 17.1, where the N−1 electrical limit was 80 MW and the airflow limit lower still.)

17.3 Local and constrained capacity

Facility-level capacity says how much IT load the building can carry in aggregate; it does not say where that load can be placed. This section moves from the facility to zones and racks, states capacity as the solution of an optimization problem under thermal limits and contingencies, lists the kinds of capacity that enter the system minimum, and shows how the electrical and thermal limits couple.

17.3.1 Local capacity below facility capacity

A facility may have 20 MW of cooling headroom in total, measured as the difference between the plant's capacity and the plant's load, while one row can accept only 0.5 MW more before its racks exceed their inlet limits. The plant has spare chillers and spare chilled water; the

row does not have spare air. Its air handlers are already at full speed, or its perforated tiles or containment cannot deliver more supply air to those positions, or its racks already receive recirculated exhaust from their neighbors. The plant headroom cannot reach the row.

The general statement is that facility capacity is not rack capacity and not zone capacity. Each zone z has its own capacity C_z, set by the cooling and airflow that can be delivered to it and by the power that can be distributed to it, and the sum of the zone capacities can exceed the facility capacity while an individual zone is full. A 100 kW rack, a dense accelerator rack for example, cannot be placed wherever there is floor space: it needs a position whose power circuits, airflow delivery and recirculation conditions support 100 kW, and in a room built for 15 kW racks there may be few or none, whatever the plant headroom.

17.3.2 Rack-level capacity

At the finest level, the margin of rack i is

$$M_i = \dot{Q}_{\mathrm{cool,max},i} - P_{\mathrm{rack},i}, \tag{17.6}$$

the largest heat the cooling can remove from that rack while its limits hold, minus the rack's power, which is equal to the heat it generates. For an illustrative rack whose cooling supports 60 kW and which draws 50 kW,

$$M_i = 60\,\mathrm{kW} - 50\,\mathrm{kW} = 10\,\mathrm{kW}.$$

The 60 kW is itself a calculated quantity. If the rack receives $\dot{m}_{\mathrm{air}} = 3.0\,\mathrm{kg/s}$ of supply air at 22°C and its exhaust must stay at or below 42°C (an illustrative hot-aisle limit), the heat the air can carry is, by $\dot{Q} = \dot{m}_{\mathrm{air}}\, c_{p,\mathrm{air}}\, \Delta T$ with $c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ/(kg\,K)}$ (Chapter 5),

$$\dot{Q}_{\mathrm{cool,max},i} = 3.0 \times 1.0 \times (42 - 22) = 60\,\mathrm{kW}.$$

The rack's capacity depends on its neighbors. If a neighboring rack is loaded more heavily and part of its exhaust recirculates into the inlet of rack i, the inlet warms. With the inlet at 24°C instead of 22°C, the same airflow carries only $3.0 \times 1.0 \times (42 - 24) = 54\,\mathrm{kW}$, and the margin of rack i falls from 10 kW to 4 kW, although nothing about rack i changed. Rack capacity is context-dependent: it is a property of the rack in its place, with its neighbors' loads and the room's airflow pattern, and the recirculation fraction α that governs it is an inferred quantity (Chapter 6).

17.3.3 Capacity under thermal limits

Collecting the rack limits into one statement turns capacity into an optimization problem. Let T_i be the inlet temperature of rack i, predicted by a model of the room as a function of the IT load P_{IT} and its distribution, the controls u (setpoints, fan speeds, staged units), the disturbances d (weather, the loads of other zones) and the parameters θ (airflow resistances, recirculation fractions, coil conductances). The thermally limited IT capacity is

$$P_{\mathrm{IT,max}} = \max\ P_{\mathrm{IT}} \quad \text{subject to} \quad T_i(P_{\mathrm{IT}}, u, d, \theta) \le T_{\max} \ \text{ for every rack } i. \tag{17.7}$$

The problem has the form of every capacity question in this chapter: the largest load such that every constraint holds. The constraints are the outputs of a physical model, not fixed numbers, so the capacity inherits the model's dependence on u, d and θ. Two consequences follow. Capacity can be increased by changing u as well as by adding equipment, a lower supply temperature or a higher fan speed raises $P_{\mathrm{IT,max}}$ at a cost in energy. And capacity is only as accurate as θ, which for airflow and recirculation is inferred from sparse temperature measurements and therefore uncertain, the subject of Section 17.4.

17.3.4 Contingency-secure IT capacity

Adding the redundancy requirement adds the contingency to the arguments of the thermal model:

$$P_{\mathrm{IT,max}}^{\mathrm{secure}} = \max\ P_{\mathrm{IT}} \quad \text{subject to} \quad T_i(P_{\mathrm{IT}}, u, d, \theta, c) \le T_{\max} \ \text{ for every rack } i \text{ and every } c \in \mathcal{C}. \tag{17.8}$$

This is the maximum contingency-secure IT load: the load the room can carry such that, if any credible contingency occurs, every rack stays within its limit. Because the constraint set of equation (17.8) contains that of equation (17.7), the secure capacity is never larger than the normal-state capacity.

As an illustration, a hall whose thermally limited capacity is 10 MW in the normal state might be limited to 8.6 MW if the air handler serving its densest rows fails, to 9.1 MW if one chilled-water pump fails and to 9.4 MW if one UPS module fails and its load moves to the remaining modules. The contingency-secure capacity is the smallest of these, 8.6 MW, set by the loss of the air handler. In practice each contingency is a separate solve of the room model with one component removed, the same procedure as screening contingencies in a power-flow study.

17.3.5 Kinds of capacity

The capacities that enter the system minimum are of several kinds: electrical distribution capacity C_E (transformers, switchgear, PDUs, branch circuits), cooling capacity C_Q (chillers, towers, pumps, coils), airflow capacity C_A (air handlers, containment, tile and duct delivery), generator capacity C_G and UPS capacity C_U, together with space and network. The system capacity is

$$C_{\mathrm{system}} = \min(C_E,\ C_Q,\ C_A,\ C_G,\ C_U,\ \dots), \tag{17.9}$$

the general form of equation (17.1). The minimum is meaningful only if all terms are converted to the same units, and the conversion is not always one to one. A UPS system carries only the IT load, so its capacity in IT terms is its rating. A generator plant carries the IT load and the cooling, so its capacity in IT terms is its rating divided by the ratio of facility power to IT power at the worst condition it must cover: a generator plant with 96 MW of firm capacity, at a facility power of 1.2 times the IT power, supports $96/1.2 = 80$ MW of IT load. The coupling between electrical and cooling limits behind that division is the subject of the next subsection.

17.3.6 Coupled electrical and thermal limits

The limits are not independent. One additional megawatt of IT load adds about one megawatt of heat, which the cooling system must remove, and removing it costs electrical power: perhaps 0.2 MW for compressors, fans and pumps. The facility's electrical demand therefore rises by about 1.2 MW for each megawatt of IT load, and the electrical limit applies to the sum. With $P_{\mathrm{cool}} = 0.2\,P_{\mathrm{IT}}$ and other overheads neglected,

$$P_{\mathrm{facility}} = P_{\mathrm{IT}} + P_{\mathrm{cool}} = 1.2\,P_{\mathrm{IT}} \le P_{\mathrm{lim}} \quad \Longrightarrow \quad P_{\mathrm{IT,max}} = \frac{P_{\mathrm{lim}}}{1.2}, \tag{17.10}$$

where P_{lim} is the facility-level electrical limit. For a utility limit $P_{\mathrm{lim}} = 100$ MW,

$$P_{\mathrm{IT,max}} = \frac{100}{1.2} = 83.3\,\mathrm{MW}.$$

After the failure of one transformer the secure facility capacity is $P_{\mathrm{lim}} = 90$ MW, and

$$P_{\mathrm{IT,max,N-1}} = \frac{90}{1.2} = 75\,\mathrm{MW}.$$

The 100 MW connection supports 83.3 MW of IT load in the normal state and 75 MW of contingency-secure IT load.

Figure 17.3 draws equation (17.10). The horizontal axis is the IT load and the vertical axis is the facility power, both in megawatts. The solid line is the facility power with the cooling overhead included; the dotted line is the IT power alone, the facility power if cooling were free. The two horizontal lines are the utility limit of 100 MW and the N−1 limit of 90 MW. Where the solid line crosses each limit, the dashed drop lines read off the IT capacity: 83.3 MW and 75 MW. The dotted line would cross the utility limit at 100 MW; the gap between 100 and 83.3 MW is the cooling power that the electrical system must also carry.

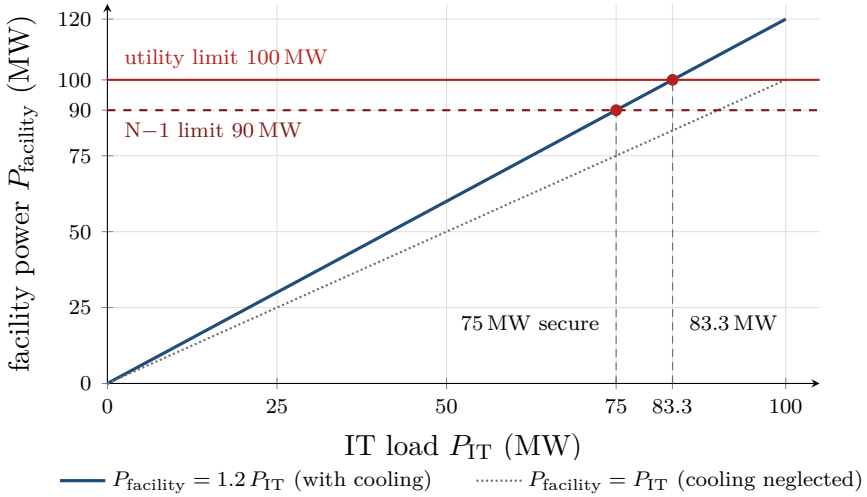


Figure 17.3: Coupled electrical and thermal limits, illustrative: $P_{\text{facility}} = P_{\text{IT}} + P_{\text{cool}}$ with $P_{\text{cool}} = 0.2\,P_{\text{IT}}$ (equation (17.10)). A 100 MW utility limit supports $100/1.2 = 83.3$ MW of IT load; the 90 MW secure limit after one transformer failure supports $90/1.2 = 75$ MW.

The factor 1.2 is the ratio of facility power to IT power, the PUE of Chapter 11, and it is not a constant. It rises in hot weather and at high load, when the chillers work against a hotter condenser, so the electrical capacity expressed in IT terms falls exactly when the cooling capacity falls too: at a peak ratio of 1.25 the 100 MW connection supports 80 MW of IT load, not 83.3 MW. Because the cooling power depends on the IT load and the weather in a nonlinear way, $P_{\text{IT,max}}$ is in general the solution of the nonlinear equation $P_{\text{IT}} + P_{\text{cool}}(P_{\text{IT}}, T_{\text{amb}}) = P_{\text{lim}}$, not a fixed division, and facility capacity is a nonlinear function of the state.

17.4 Probabilistic capacity

Every capacity so far was a number. The model behind each one has uncertain inputs, so each capacity is in fact a distribution. This section shows why a deterministic limit can mislead, defines capacity through a chance constraint, lists the sources of uncertainty, computes the probability of a negative margin, and introduces the quantile capacities that express the result.

17.4.1 Deterministic limits and the probability of violation

Suppose a model predicts that the hottest rack inlet will reach 29.8°C at the planned load, against a limit $T_{\max} = 30$°C. A deterministic reading says the plan is safe: $29.8 < 30$. Suppose also that the prediction is uncertain, because the rack airflows and the recirculation fractions behind it are inferred rather than measured, and that the uncertainty is normal with a standard deviation $\sigma_T = 2$ K. The temperature is then a random variable $T \sim \mathcal{N}(29.8, 2^2)$, and the

probability that it exceeds the limit is

$$\Pr(T > T_{\max}) = \Pr\!\left(Z > \frac{30 - 29.8}{2}\right) = 1 - \Phi(0.1) = 1 - 0.540 = 0.460,$$

where Z is a standard normal variable and Φ its cumulative distribution function. The "safe" plan has a 46 % probability of violating the limit. A mean below the limit is not a low risk; when the uncertainty is large compared with the distance to the limit, the probability of violation is close to one half.

Principle 17.2. *A predicted value below a limit does not mean a low probability of exceeding it. The risk is set by the distance to the limit measured in standard deviations of the prediction, not in degrees or megawatts.*

17.4.2 Chance-constrained capacity

The remedy is to state the limit as a probability. Instead of requiring $T_i \le T_{\max}$, require

$$\Pr(T_i > T_{\max}) \le \epsilon \quad \text{for every rack } i, \tag{17.11}$$

with a risk tolerance ϵ set by policy, for example $\epsilon = 0.01$. The capacity becomes the largest load that satisfies the chance constraint,

$$C_{99} = \max\ P_{\mathrm{IT}} \quad \text{subject to} \quad \Pr\!\big(T_i(P_{\mathrm{IT}}, u, d, \theta) > T_{\max}\big) < 0.01 \ \text{ for every rack } i, \tag{17.12}$$

where the probability is taken over the uncertainty in θ, d and the load itself. The subscript records the confidence, $1 - \epsilon$, in percent. The result is a risk-based capacity rating: it answers "how much load can be carried with a violation probability below 1 %" rather than "how much load can be carried if every input takes its nominal value".

The same idea applies to any margin. With capacity and load both uncertain, the general risk-constrained capacity is

$$C_\epsilon = \max\ L \quad \text{subject to} \quad \Pr(M < 0) < \epsilon, \qquad M = C - L. \tag{17.13}$$

17.4.3 A deterministic and a risk-based rating

Consider an illustrative hall whose deterministic rating is $P_{\mathrm{IT,max}} = 10\,\mathrm{MW}$. At 10 MW the probability that the hottest inlet exceeds its limit is 12 %; at 9.2 MW it is 1 %. The two ratings are

$$C_{\mathrm{det}} = 10\,\mathrm{MW}, \qquad C_{99} = 9.2\,\mathrm{MW},$$

and the difference, 0.8 MW, is a risk margin: the load that must be withheld to bring the violation probability from 12 % down to 1 %.

The numbers can be reproduced with a simple illustrative model of the hottest inlet temperature. Let its predicted mean rise linearly with the IT load, $\mu_T(P_{\mathrm{IT}}) = 29.3 + 0.87\,(P_{\mathrm{IT}} - 10)$°C, with a prediction standard deviation $\sigma_T = 0.6\,\mathrm{K}$ and $T_{\max} = 30$°C. The deterministic rating is the load at which the predicted mean sits a fixed design allowance of 0.7 K below the limit, $\mu_T = 29.3$°C, which is 10 MW. At that load

$$\Pr(T > T_{\max}) = 1 - \Phi\!\left(\frac{30 - 29.3}{0.6}\right) = 1 - \Phi(1.17) = 1 - 0.879 = 0.121 \approx 12\,\%.$$

The 1 % capacity requires the mean to sit $z_{0.99}\,\sigma_T$ below the limit, with $z_{0.99} = 2.326$ the 99th percentile of the standard normal: $\mu_T = 30 - 2.326 \times 0.6 = 28.60$°C, which the model reaches at

$$P_{\mathrm{IT}} = 10 - \frac{29.3 - 28.60}{0.87} = 10 - 0.80 = 9.2\,\mathrm{MW}.$$

Without the allowance, a deterministic rating that put the mean exactly at the limit would sit at $10 + 0.7/0.87 = 10.8\,\text{MW}$, where the violation probability is 50 %. A fixed allowance is a crude chance constraint: it buys a definite but unstated violation probability, here 12 %, and the probability it buys changes whenever σ_T changes.

Figure 17.4 draws the violation probability of this model against the IT load. The horizontal axis is the IT load; the vertical axis is the probability that the hottest inlet exceeds 30°C, on a logarithmic scale so that 1 % and 50 % are both legible. The curve rises steeply: from about 0.04 % at 8.5 MW to 50 % at 10.8 MW. The dotted horizontal line is the tolerance $\epsilon = 1\,\%$. The two marked points are the risk-based rating at 9.2 MW and the deterministic rating at 10 MW, and the bracket between them is the 0.8 MW risk margin. The steepness is the useful lesson of the figure: near the rating, a few hundred kilowatts move the violation probability by an order of magnitude.

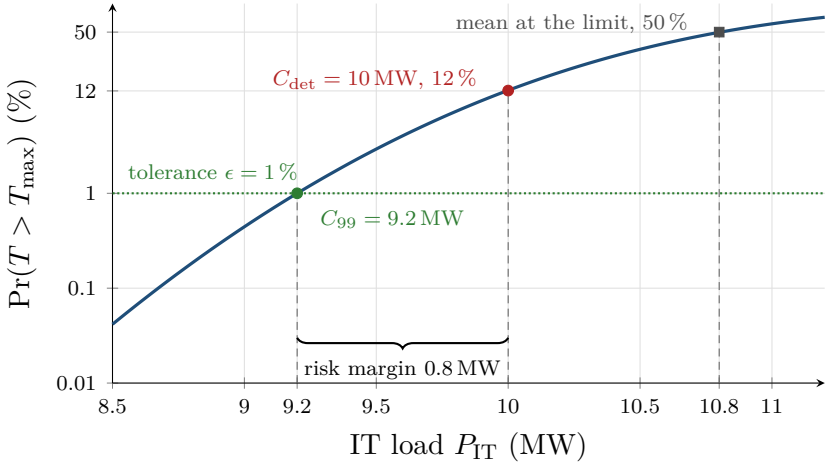


Figure 17.4: Probability that the hottest inlet exceeds $T_{\max} = 30°\text{C}$ against IT load, illustrative: $\Pr(T > T_{\max}) = 1 - \Phi((30 - \mu_T)/\sigma_T)$ with $\mu_T = 29.3 + 0.87\,(P_{\text{IT}} - 10)°\text{C}$ and $\sigma_T = 0.6\,\text{K}$ (logarithmic vertical axis). The deterministic rating of 10 MW carries a 12 % violation probability; the 1 % rating is 9.2 MW.

17.4.4 Sources of uncertainty

The uncertainty in a capacity comes from several places, and each enters the model through a different variable. Table 17.1 lists the main ones, whether each is measured, forecast or inferred, and which side of the margin it affects. Rack and aisle airflows and the coil conductance UA are rarely measured at the resolution the model needs and are inferred from temperatures, as Chapters 14 and 15 develop. The IT load and the outdoor temperature are measured now and forecast for later. The state of each piece of equipment is reported by the control system but is itself uncertain over a planning horizon, because units can fail. And every measurement carries a possible bias b: an inlet sensor that reads 1 K low hides 1 K of margin. All of these propagate through the model into a distribution of the available capacity, $p(C)$, and of the load, $p(L)$.

17.4.5 The margin as a distribution

With capacity and load both random, $C \sim p(C)$ and $L \sim p(L)$, the margin $M = C - L$ is random as well, and the natural risk metric is the probability that it is negative, $\Pr(M < 0)$, the probability that the load exceeds the capacity. If C and L are independent and normal,

Table 17.1: Sources of uncertainty in a capacity margin. "Side" says whether the source acts mainly on the capacity C or on the load L.

Source	Kind	Side	How it enters
Rack and aisle airflow $\dot{m}_{\text{air}}$	inferred	C	sets the heat each rack's air can carry and the recirculation into its inlet
Coil conductance UA	inferred	C	sets the heat a coil can transfer; falls with fouling
IT power forecast	forecast	L	future heat and electrical load
Outdoor temperature forecast	forecast	C, L	chiller and tower capacity; cooling power in the facility load
Equipment state	reported, uncertain ahead	C	which units are available; failures remove capacity
Sensor bias b	inferred	C	shifts the measured margin to a temperature limit

$C \sim \mathcal{N}(\mu_C, \sigma_C^2)$ and $L \sim \mathcal{N}(\mu_L, \sigma_L^2)$, then M is normal with

$$\mathbb{E}[M] = \mu_C - \mu_L, \qquad \sigma_M^2 = \sigma_C^2 + \sigma_L^2, \qquad \Pr(M < 0) = \Phi\!\left(-\frac{\mathbb{E}[M]}{\sigma_M}\right). \tag{17.14}$$

The ratio $\mathbb{E}[M]/\sigma_M$ is the distance from the expected margin to zero measured in standard deviations, and it alone fixes the risk.

For an illustrative cooling capacity $C \sim \mathcal{N}(10, 0.8^2)\,\text{MW}$ and an illustrative heat load $L \sim \mathcal{N}(8.5, 0.5^2)\,\text{MW}$,

$$\mathbb{E}[M] = 10 - 8.5 = 1.5\,\text{MW}, \qquad \sigma_M^2 = 0.8^2 + 0.5^2 = 0.64 + 0.25 = 0.89\,\text{MW}^2,$$

so $\sigma_M = \sqrt{0.89} = 0.943\,\text{MW}$. The expected margin is $1.5/0.943 = 1.59$ standard deviations from zero, and

$$\Pr(M < 0) = \Phi(-1.59) = 1 - \Phi(1.59) = 1 - 0.944 = 0.056.$$

A margin of 1.5 MW, which looks comfortable, carries a 5.6 % probability that load exceeds capacity: roughly one hour in eighteen if the conditions persisted. The overload risk is not negligible.

Figure 17.5 draws the example. In the left panel the horizontal axis is power in megawatts and the vertical axis is probability density; the dashed curve is the load distribution and the solid curve the capacity distribution. The two overlap: there are values of the load above values of the capacity. The right panel is the distribution of their difference, the margin. Its mean is 1.5 MW, its standard deviation 0.943 MW, and the red area to the left of zero is $\Pr(M < 0) = 5.6\,\%$.

When C and L are not independent, the variance of the margin carries a covariance term, $\sigma_M^2 = \sigma_C^2 + \sigma_L^2 - 2\,\text{Cov}(C, L)$. A covariance that is negative, capacity tending to be low when load is high, increases σ_M and the risk; Section 17.5.2 shows that weather produces exactly that sign. When the distributions are not normal, the probability is computed by propagating samples of the uncertain inputs through the model and counting the fraction of samples with $M < 0$.

17.4.6 Quantile capacities

A distribution of capacity is summarized by its quantiles. Define C_q as the capacity that is exceeded with probability $q/100$,

$$\Pr(C \geq C_q) = \frac{q}{100}. \tag{17.15}$$

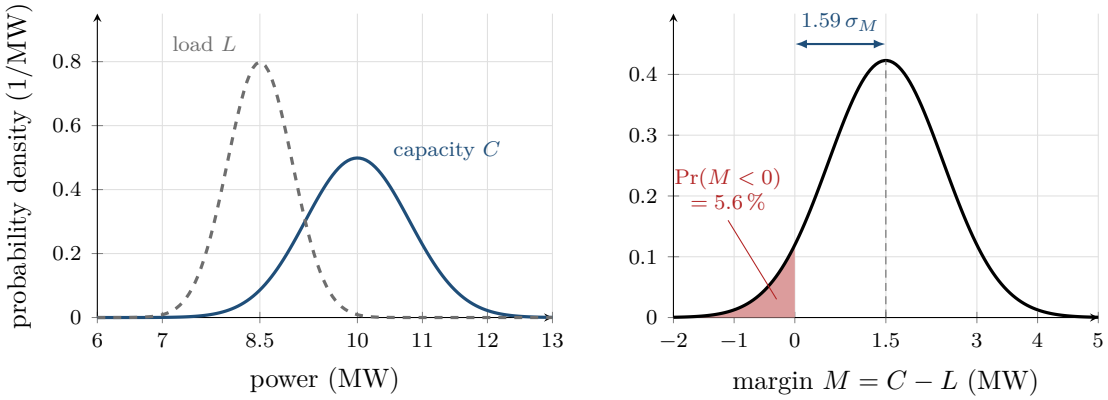


Figure 17.5: The margin as a distribution, illustrative. Left: $C \sim \mathcal{N}(10, 0.8^2)$ and $L \sim \mathcal{N}(8.5, 0.5^2)$ MW, independent. Right: $M = C - L \sim \mathcal{N}(1.5, 0.943^2)$ MW (equation (17.14)); the shaded tail is $\Pr(M < 0) = \Phi(-1.59) = 0.056$.

C_{50} is the median capacity; C_{95} the value the capacity exceeds with 95 % probability; C_{99} the value it exceeds with 99 % probability. For a load known exactly, C_q is the risk-constrained capacity of equation (17.13) with $\epsilon = 1 - q/100$: a planned load equal to C_{95} has a 5 % probability of exceeding the capacity.

For an illustrative facility, $C_{50} = 90\,\text{MW}$, $C_{95} = 82\,\text{MW}$ and $C_{99} = 78\,\text{MW}$. The three numbers answer three different questions. The median is the capacity in a typical realization of the uncertain inputs. C_{95} is a capacity for planning with a 5 % tolerance; C_{99} for a stricter policy. Which one a facility sells is a policy decision about risk tolerance, the same decision that probabilistic ratings of generation and transmission equipment encode elsewhere in energy systems. The quantiles also show the shape of the uncertainty: the gap from median to C_{95} is 8 MW and from C_{95} to C_{99} a further 4 MW, a lower tail longer than a normal distribution with the same median and C_{95} would have, for which C_{99} would be $90 - 2.326 \times 8/1.645 = 78.7\,\text{MW}$. A heavier lower tail is the expected shape for capacity: equipment outages and unfavorable combinations of inferred parameters remove capacity in large steps, while nothing adds capacity beyond what is installed.

Figure 17.6 draws a distribution with these three quantiles. The horizontal axis is the facility capacity in megawatts of IT load; the vertical axis is probability density. The curve is skewed to the left, with its peak near 91 MW and a long tail toward low capacity. The light red area to the left of C_{95} holds 5 % of the probability and the darker area to the left of C_{99} holds 1 %; the three vertical lines mark the median and the two quantile capacities.

17.5 Capacity margin in time

The margin of interest is rarely the present one; an operator deciding whether to accept a deployment, or whether to shed load before the afternoon, needs the margin hours or days ahead. This section adds load and weather forecasts to the margin, works a three-hour example, follows the slow erosion of capacity by degradation, and separates the probability of a capacity violation from the probability of equipment failure and from the consequence of a violation.

17.5.1 Load forecasts

An illustrative facility carries 8 MW of IT load against a capacity of 10 MW: a present margin of 2 MW. The IT load is forecast to reach $P_{\text{IT}}(t + 2\,\text{h}) = 9.7 \pm 0.8\,\text{MW}$ two hours from now, with the ± 0.8 MW read as one standard deviation of the forecast error. The forecast margin

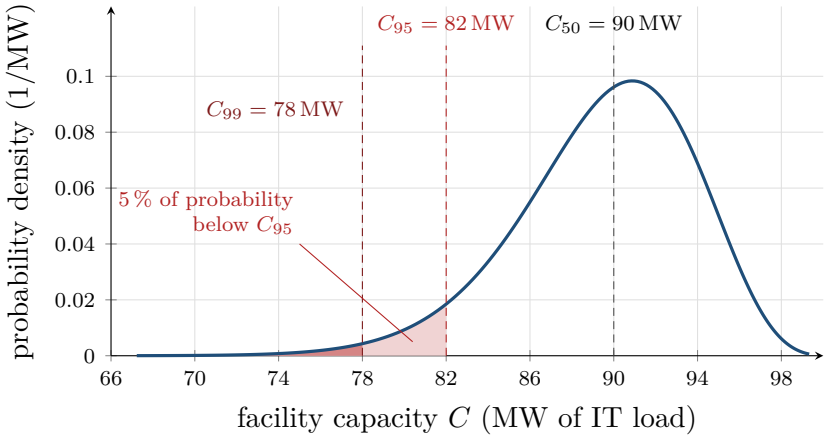


Figure 17.6: An illustrative capacity distribution with $C_{50} = 90$, $C_{95} = 82$ and $C_{99} = 78\,\text{MW}$, built from the quantile function $C = 90 + 4.15\,z - 0.433\,z^2$ MW with z standard normal; its density is $\varphi(z)/(4.15 - 0.866\,z)$ along the curve. The lower tail is longer than the upper one.

has a mean of $10 - 9.7 = 0.3\,\text{MW}$, and if the capacity stays at 10 MW,

$$\Pr\big(L(t + 2\,\text{h}) > 10\big) = 1 - \Phi\left(\frac{10 - 9.7}{0.8}\right) = 1 - \Phi(0.375) = 1 - 0.646 = 0.354.$$

The present margin is 2 MW and certain; the margin in two hours has a mean of 0.3 MW and a 35 % probability of being negative. Both numbers belong in a capacity statement: the current margin, which is measured, and the forecast margin with its uncertainty, which is predicted. In general the horizon h enters as $M(t + h) = C(t + h) - L(t + h)$, and both terms are forecasts.

17.5.2 Weather forecasts

Weather acts on both terms at once. A hot afternoon lowers the cooling capacity $C_{\text{cool}}(t + h)$, through the condenser temperatures of Section 17.2.2, and at the same time it can raise the load: server fans speed up at warmer inlets, the cooling power that the electrical system carries increases, and in many facilities the compute demand itself peaks in the afternoon. Capacity falls while load rises, and the margin shrinks from both sides.

For the probability, the important consequence is the sign of the correlation. A forecast error in the outdoor temperature moves the capacity and the load in opposite directions: if the afternoon is hotter than forecast, capacity is lower and load is higher than forecast. The errors of C and L are therefore negatively correlated, the covariance term of Section 17.4.5 is negative, and the variance of the margin is larger than the sum of the two variances. Treating the two forecasts as independent understates the risk of exactly the hours in which the risk is highest.

17.5.3 A three-hour-ahead margin

At 14:00 an illustrative facility has a cooling capacity $C_{\text{cool}} = 12\,\text{MW}$ and a heat load $L = 8\,\text{MW}$: a margin of 4 MW. The forecast for 17:00 gives $C_{\text{cool}} = 9.5\,\text{MW}$, because of the weather, and $L = 9.2\,\text{MW}$, because of compute demand: a mean margin of

$$\mathbb{E}[M(17{:}00)] = 9.5 - 9.2 = 0.3\,\text{MW}.$$

The deterministic forecast says the facility will remain within its capacity. The question that matters is $\Pr(M(t + h) < 0)$, and it requires the uncertainty of the two forecasts. With

illustrative standard deviations at the three-hour horizon of $\sigma_C = 0.5\,\mathrm{MW}$ for the capacity (from the temperature forecast) and $\sigma_L = 0.6\,\mathrm{MW}$ for the load, independent errors give

$$\sigma_M = \sqrt{0.5^2 + 0.6^2} = \sqrt{0.61} = 0.781\,\mathrm{MW}$$

and $\Pr(M < 0) = \Phi(-0.3/0.781) = \Phi(-0.384) = 0.350$. With the negative correlation of Section 17.5.2, say a correlation coefficient $\rho = -0.5$ between the two errors,

$$\sigma_M^2 = 0.25 + 0.36 - 2 \times (-0.5) \times 0.5 \times 0.6 = 0.91\,\mathrm{MW}^2, \qquad \sigma_M = 0.954\,\mathrm{MW},$$

and $\Pr(M < 0) = \Phi(-0.3/0.954) = \Phi(-0.314) = 0.377$. A forecast margin of 0.3 MW is a 35 to 38 % chance of running out of cooling at 17:00. The decision at 14:00, to defer a deployment, to pre-cool, or to arrange to shift flexible load, is made on that probability, not on the deterministic 0.3 MW.

Figure 17.7 draws the example. In the upper panel the horizontal axis is the time of day and the vertical axis is power in megawatts. The blue curve is the cooling capacity and the gray curve the heat load; both are known up to 14:00 (solid) and forecast after it (dashed), with shaded bands of plus and minus one standard deviation that widen with the horizon. The vertical dotted line marks the issue time of the forecast. The capacity curve falls and the load curve rises, and by 17:00 their bands overlap. The lower panel, on the same time axis, is the probability of a negative margin, computed at each time from the means and standard deviations of the upper panel, for independent errors (solid) and for $\rho = -0.5$ (dashed). It is negligible until about 15:30, crosses 1 % near 16:00 and reaches 35–38 % at 17:00.

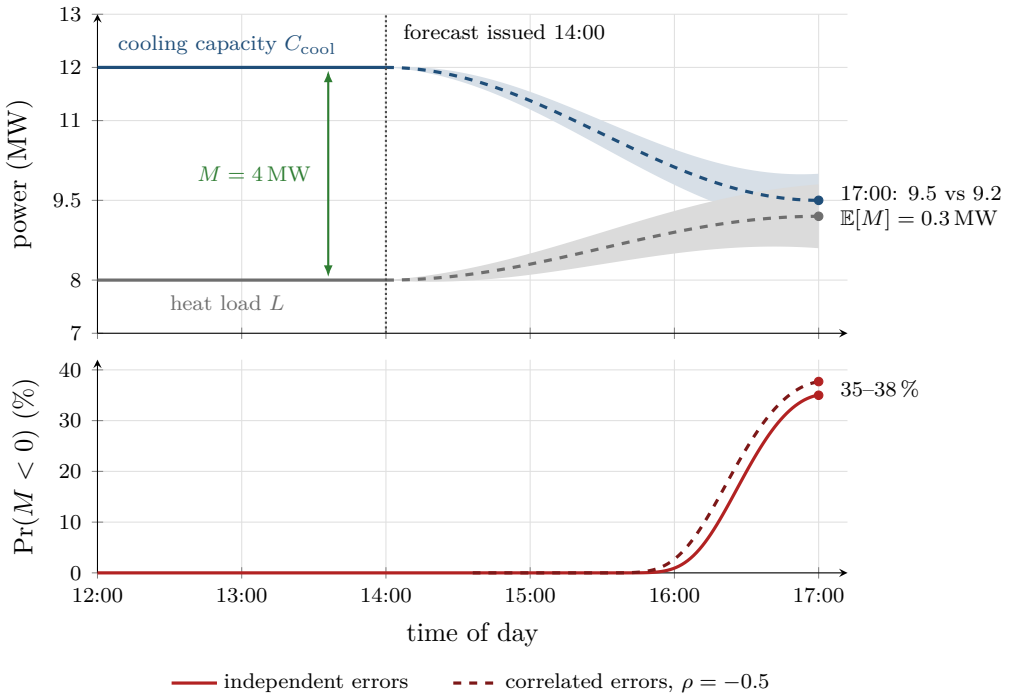


Figure 17.7: A three-hour-ahead margin, illustrative. Upper panel: cooling capacity $C_{\mathrm{cool}}(t) = 12 - 2.5\,s(t)$ and heat load $L(t) = 8 + 1.2\,s(t)$ MW, with $s(t) = (1 - \cos(\pi(t-14)/3))/2$ after the 14:00 forecast, and $\pm 1\sigma$ bands growing linearly to $\sigma_C = 0.5$ and $\sigma_L = 0.6\,\mathrm{MW}$ at 17:00. Lower panel: $\Pr(M < 0) = \Phi(-\mathbb{E}[M]/\sigma_M)$ for independent errors and for a correlation $\rho = -0.5$.

17.5.4 Gradual degradation

Not every loss of capacity is a failure. Coil fouling, scale in condenser tubes, clogged filters and worn pump impellers lower capacity slowly. Fouling, for example, lowers the conductance UA of

a cooling coil, and with it the maximum heat the coil can transfer, $\dot{Q}_{\max}(t)$, a little each month. No alarm sounds, because the control system compensates: valves open further, fans speed up, and the temperatures the alarms watch stay at their setpoints. The first alarm comes only when the compensation runs out, which is the moment the margin has already reached zero.

An illustrative cooling system with $C = 10\,\mathrm{MW}$ against a peak load of $8.5\,\mathrm{MW}$ loses $1\,\%$ of its capacity per month to fouling. After twelve months its capacity is $10 \times 0.99^{12} = 8.86\,\mathrm{MW}$ and the margin $0.36\,\mathrm{MW}$; after sixteen months it is $10 \times 0.99^{16} = 8.51\,\mathrm{MW}$ and the margin has effectively vanished. Throughout those months every temperature reading was normal. The first hot afternoon that lowers capacity by a further megawatt, as in Section 17.2.2, produces a violation well before the sixteenth month. This is the argument for tracking the margin itself, computed from the inferred UA and the measured duty, rather than waiting for the temperatures: the margin shows a steady decline months before the temperature alarm does, and the decline can be scheduled away by cleaning.

17.5.5 Capacity risk and equipment failure probability

The probability that equipment fails and the probability that capacity is violated are different quantities. A facility can have reliable equipment and a high capacity risk, if it runs with little margin; it can have unreliable equipment and a low capacity risk, if it holds enough redundancy.

Two illustrative facilities each serve a peak heat load $L \sim \mathcal{N}(9.0, 0.4^2)\,\mathrm{MW}$ with $5\,\mathrm{MW}$ chillers. Facility X has two chillers, each unavailable at a given peak hour with probability 0.005. Facility Y has three older chillers, each unavailable with probability 0.05, ten times as often. In X, both chillers are available with probability $0.995^2 = 0.990$, giving $C = 10\,\mathrm{MW}$ and an overload probability $\Pr(L > 10) = 1 - \Phi(2.5) = 0.0062$; if either is out, $C = 5\,\mathrm{MW}$ and overload is certain. In Y, all three are available with probability 0.857 ($C = 15\,\mathrm{MW}$, overload negligible), exactly one is out with probability $3 \times 0.05 \times 0.95^2 = 0.135$ ($C = 10\,\mathrm{MW}$, overload probability 0.0062), and two or more are out with probability 0.0073 (overload certain). The overload probabilities are

$$\text{X: } 0.990 \times 0.0062 + 0.010 \times 1 = 0.016, \qquad \text{Y: } 0.135 \times 0.0062 + 0.0073 \times 1 = 0.008.$$

Y's equipment fails ten times as often and Y's capacity risk is half of X's. Reliability of the parts, which Chapter 12 quantifies, is one input to capacity risk; the margin and the redundancy decide how much of it becomes a violation.

Principle 17.3. *Equipment failure probability is not capacity violation probability. A violation needs load to exceed the capacity that remains, so a facility with reliable equipment and a thin margin can carry more capacity risk than one with unreliable equipment and ample redundancy.*

17.5.6 Consequence and expected risk

Two zones can share the same violation probability and differ completely in what a violation costs. Suppose each of two zones has $\Pr(M < 0) = 5\,\%$ over the coming peak season. One hosts noncritical batch computing, whose jobs can be paused and rerun; the other hosts a critical service with contractual availability. The expected risk weights the probability by the consequence ℓ of a violation,

$$\mathcal{R} = \Pr(M < 0) \times \ell. \tag{17.16}$$

With an illustrative consequence of $\ell = \$20\,000$ for the batch zone (rerun jobs, a delayed deadline) and $\ell = \$5$ million for the critical zone (breached service agreements, lost customers),

$$\mathcal{R}_{\text{batch}} = 0.05 \times \$20\,000 = \$1\,000, \qquad \mathcal{R}_{\text{critical}} = 0.05 \times \$5\,000\,000 = \$250\,000.$$

The same probability carries a risk 250 times larger in the critical zone. Capacity decisions, where to place new load, which zone gets the next cooling upgrade, which load to shed first, are made on $\mathcal{R}$, and the same probability supports a different decision in each zone. A natural consequence is to give each zone its own tolerance ϵ, looser where violations are cheap.

17.6 Bottlenecks and capacity planning

The computations above produce a capacity, a margin and a risk; planning asks what to do with them. This section quantifies stranded capacity, attributes the usable capacity to its bottleneck, defines the marginal value of an investment, propagates a proposed deployment through the facility, draws capacity as a region in the space of operating conditions, and assembles the result into the report an operator can use.

17.6.1 Stranded capacity

Stranded capacity, as introduced in Chapter 11, is installed capacity that the binding limit prevents from being used. In an illustrative facility with 100 MW of electrical capacity and 85 MW of cooling capacity, 15 MW of electrical capacity is stranded: paid for, maintained, and unusable until the cooling is expanded. If, in addition, poor airflow in certain halls limits the IT load that can actually be deployed to 78 MW, a further $85 - 78 = 7$ MW is stranded, this time spatially: the plant can cool 85 MW, but not in the places where the racks can go. In total, $100 - 78 = 22$ MW of electrical capacity, more than a fifth, is stranded.

Stranded capacity is a cost as well as a constraint. Electrical and cooling infrastructure is commonly priced in the millions of dollars per megawatt, so stranded megawatts are capital that earns nothing. They are also a planning signal: they say which subsystem is oversized relative to the bottleneck, and that a modest investment at the bottleneck may unlock a large amount of already-paid capacity.

17.6.2 Capacity attribution

Attribution states which limit sets the usable capacity and by how much each other limit exceeds it. Table 17.2 works an illustrative facility, a second example with the same 100 MW of electrical capacity and the same 78 MW of usable capacity as the one above, but with its cooling split into the chilled-water plant and the air handlers (CRAH units) that deliver the cooling to the room. The limits are electrical 100 MW, chilled-water plant 90 MW, CRAH capacity 84 MW and local airflow 78 MW. The usable capacity is 78 MW, and the active bottleneck is the local airflow. The table also lists, for each limit, its excess over the usable capacity and the usable capacity gained by adding 1 MW to that limit alone.

Table 17.2: Capacity attribution for an illustrative facility. All values in megawatts of IT load. The last column is the change in usable capacity from adding 1 MW to that limit alone.

Limit	Capacity	Excess over usable	Status	+1 MW gives
Electrical distribution	100	22	stranded	0
Chilled-water plant	90	12	stranded	0
CRAH capacity	84	6	next to bind	0
Local airflow	78	0	active bottleneck	+1
Usable capacity	78			

The table answers the investment question directly. Improving the local airflow, with containment, blanking panels, tile changes or ducting, is where investment has value: each megawatt of airflow capacity adds a megawatt of usable capacity, up to 6 MW, when the CRAH capacity of 84 MW becomes the bottleneck. Beyond that, the next 6 MW requires CRAH capacity, and only beyond 90 MW does the chilled-water plant matter. An investment in the electrical system, or in the chiller plant, adds nothing until the limits below it have been raised.

17.6.3 Marginal usable capacity

The last column of Table 17.2 is a derivative. The marginal usable capacity of limit j is the change in usable capacity per unit change of that limit,

$$\frac{\Delta C_{\text{usable}}}{\Delta C_j} = \begin{cases} 1 & \text{if } j \text{ alone binds and } \Delta C_j \text{ is below the gap to the next limit,} \\ 0 & \text{if } j \text{ does not bind.} \end{cases} \tag{17.17}$$

For a minimum of limits, the derivative is one for the binding term and zero for all others. An illustrative new CRAH unit adds 1 MW of local cooling capacity. If the electrical system is the bottleneck at 80 MW, it adds 0 MW of usable capacity: the room can cool more, but the extra cooling has no load to cool that the electrical system can supply. The same unit installed in a facility whose bottleneck is the local cooling of that room adds its full 1 MW.

Capacity planning should therefore be based on the marginal usable capacity of each candidate investment, not on the nameplate capacity it adds. When the limits are uncertain, the derivative becomes an expected value: an investment at a limit that binds with probability 0.6 adds, in expectation, about 0.6 MW of usable capacity per megawatt, and a limit that never binds adds nothing however large it is.

> **Principle 17.4.** *An investment adds usable capacity only at the binding limit. Plan on the marginal usable capacity of each option, which is zero for every limit that does not bind.*

17.6.4 What-if capacity analysis

A static rack-power limit, "no rack above 20 kW" or "no row above 1 MW", is a proxy for the physics. The direct alternative is to propagate a proposed change through the coupled systems and check every limit along the way. Consider adding 500 kW of IT load to row 4. The chain of consequences is

$$P_{\text{IT}} \uparrow \Rightarrow\ \dot{Q}_{\text{rack}} \uparrow \Rightarrow\ T_{\text{rack}} \uparrow \Rightarrow\ \dot{Q}_{\text{CHW}} \uparrow \Rightarrow\ P_{\text{cool}} \uparrow \Rightarrow\ P_{\text{facility}} \uparrow,$$

where $\dot{Q}_{\text{CHW}}$ is the heat carried by the chilled-water loop, and each link has its own limit.

Figure 17.8 draws the chain. Each box on the left is a quantity that the change moves, from the IT load at the top to the facility power at the bottom; the arrows between them are the physical relations. Each box on the right is the limit checked at that step, with the illustrative result. With illustrative numbers: row 4 carries 800 kW and receives at most 90 kg/s of supply air at 22°C. After the change it carries 1300 kW, which at a 12 K rise across the racks needs $1300/(1.0 \times 12) = 108$ kg/s of air. The 18 kg/s shortfall is made up by recirculated exhaust, a recirculation fraction $\alpha = 18/108 = 0.17$, and by the mixing rule of Chapter 6, $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{out}}$ with $T_{\text{out}} = T_{\text{in}} + 12$ K, the row's average inlet becomes $22 + 12 \times 0.17/0.83 = 24.4$°C; the hottest rack, nearest the end of the row where recirculation concentrates, is warmer. The chilled-water load rises by about 0.5 MW, the cooling power by about 0.1 MW, and the facility power by about 0.6 MW, each to be checked against the plant

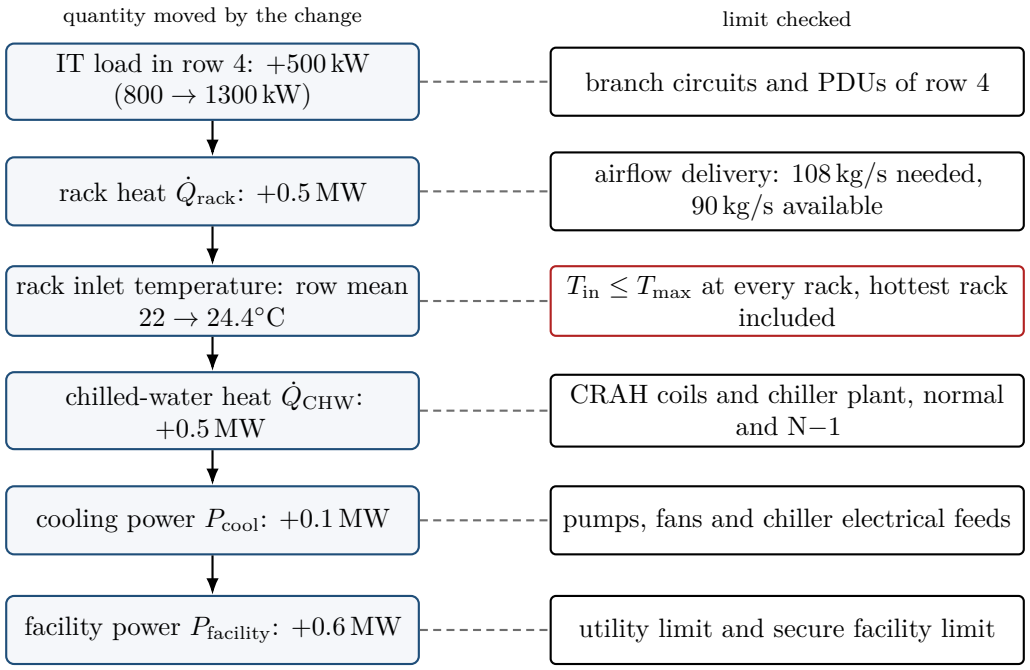


Figure 17.8: What-if capacity analysis of adding 500 kW to row 4, illustrative. The left column propagates the change from IT load to facility power; the right column is the limit checked at each step. The airflow shortfall raises the row's mean inlet by recirculation, and the inlet limit at the hottest rack (red) is the one most likely to fail.

and the utility margins and their N−1 versions. If every check passes, the deployment is feasible; if one fails, the chain names the limit that fails.

The analysis is more informative than a static limit in both directions. A row with good airflow can accept more than the static limit allows, and a row with poor airflow can fail at a load the static limit permits. Chapter 18 develops what-if analysis as a method; here it is one way of computing a capacity, by asking whether a specific increment is feasible.

17.6.5 Capacity as a feasible region

Because capacity depends on the disturbances, the parameters and the state, $C = f(d, \theta, \text{state})$, it is not a number but a surface over the space of conditions. In the plane of IT load and outdoor temperature, the facility is safe in a region bounded by that surface: an illustrative facility supports $P_{\text{IT,max}} = 90$ MW in mild weather and 75 MW in extreme weather, with the boundary falling between. An operating point $(P_{\text{IT}}, T_{\text{amb}})$ inside the region is feasible; outside it, some limit is violated.

With uncertainty, the boundary is not a line but a band. For a given operating point the question becomes the probability that capacity is below the load, and an operator can state either a probability, such as $\Pr(C < 80\,\text{MW}) = 0.2$ at a warm condition, or a quantile capacity, such as C_{95}. Figure 17.9 draws such a fuzzy boundary. The horizontal axis is the outdoor temperature and the vertical axis is the IT load. The solid curve is the median capacity C_{50}, flat at 90 MW up to 20°C and falling linearly to 75 MW at 40°C. The dashed curve is C_{95} and the dotted curve C_{99}, 8 and 12 MW below the median, using at every temperature the skewed distribution of Figure 17.6. The green region below C_{99} is feasible with at least 99 % probability; the gold band between C_{99} and C_{50} is the fuzzy boundary; above C_{50}, a violation is more likely than not. The horizontal line is a planned IT load of 80 MW. At 15°C it lies between C_{99} and C_{95}, with a violation probability of 2.3 %; at 28.3°C, where the median capacity is 83.8 MW,

the probability is 20 %, the $\Pr(C < 80) = 0.2$ of the example; at 40°C the median is 75 MW and the probability is 92 %.

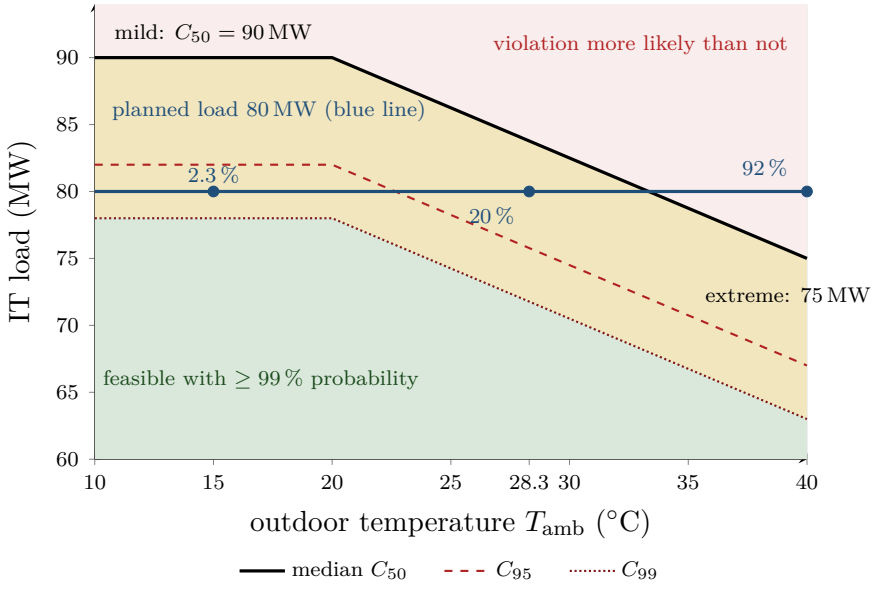


Figure 17.9: Capacity as a feasible region in the plane of outdoor temperature and IT load, illustrative: $C_{50}(T_{amb}) = 90$ MW up to 20°C, then $90 - 0.75\,(T_{amb} - 20)$ MW; at each temperature $C = C_{50} + 4.15\,z - 0.433\,z^2$ with z standard normal, so $C_{95} = C_{50} - 8$ and $C_{99} = C_{50} - 12$ MW. The marked probabilities are $\Pr(C < 80\,\text{MW})$ at 15, 28.3 and 40°C.

17.6.6 Repeated calculation

All of the preceding pieces combine into one calculation that is physics-constrained (the limits are outputs of electrical, airflow and thermal models), uncertain in its states and parameters (airflows, recirculation, coil conductance, sensor biases) and dependent on changing conditions (weather, load, equipment state). The question it answers is

how much additional IT load can be added safely right now?

and the answer has a short shelf life. It must be recomputed when the weather changes, because C_{cool} follows the outdoor conditions; when equipment goes out of service or returns, because the contingency set and the available capacity change; when fouling progresses, because UA and with it $\dot{Q}_{max}$ fall; and when the sensor evidence changes, because a new measurement narrows or shifts the distribution of an inferred parameter and with it $p(C)$. Capacity in this sense is a monitored quantity, like a temperature, and not a design constant, like a nameplate.

17.6.7 A practical capacity report

A capacity statement that an operator can act on contains more than one number. Table 17.3 lists the items of an illustrative report, with the kind of quantity each is. It gives the additional IT capacity available now at a stated confidence, 6.2 MW at 95 %; the bottleneck that sets it, the cooling airflow in hall B; the additional capacity under the N−1 requirement, 2.8 MW; a forecast for the next stressful condition, 1.5 MW tomorrow at 16:00 because of a high wet-bulb temperature; the uncertainty that dominates the result, the rack airflow estimate; and the measurement that would reduce that uncertainty most, a direct airflow measurement in rows 7 to 9.

Table 17.3: An illustrative capacity report. Each item names the quantity, its value, and whether it is calculated, forecast or inferred.

Item	Value	Kind
Additional IT capacity now	6.2 MW at 95 % confidence	calculated from inferred parameters
Primary bottleneck	cooling airflow in hall B	attribution of the minimum
Additional capacity, N−1	2.8 MW	calculated over the contingency set
Forecast, tomorrow 16:00	1.5 MW, high wet-bulb temperature	forecast of weather and load
Dominant uncertainty	rack airflow estimate	inferred, largest contribution to σ_C
Most valuable measurement	direct airflow, rows 7 to 9	chosen for the uncertainty it removes

Each line answers a decision. The first says how much can be sold or deployed today. The second says where an investment would help, by the marginal usable capacity of Section 17.6.3. The third says how much of the first survives a failure. The fourth warns of the condition in which the margin will be smallest. The last two connect the capacity statement to the inference behind it: the 6.2 MW is uncertain mainly because of one inferred quantity, and one measurement would sharpen it. Choosing that measurement is the subject of Chapter 19.

17.7 What to retain

Definition 17.1 (Capacity under risk)**.** Nameplate capacity, usable capacity, secure capacity and risk-constrained capacity are four different quantities, in decreasing order. The capacity of a data center is the additional load the system can support under current conditions, with every subsystem limit and required contingency respected, while the probability of a violation, $\Pr(M < 0)$ with $M = C - L$, stays below an accepted tolerance.

The equations to keep are the margin $M = C - L$, its risk $\Pr(M < 0)$, the system minimum $C_{\text{system}} = \min_i C_i$ and the risk-constrained capacity $C_\epsilon = \max L$ subject to $\Pr(M < 0) < \epsilon$. Table 17.4 lists the chapter's objects and the subsection that defines each.

17.8 Exercises

Exercise 17.1. A facility has the following limits, all in megawatts of IT load: electrical distribution 60, chilled-water plant 52, CRAH capacity 49, local airflow 46. (a) Compute the usable capacity, name the bottleneck, and compute the stranded capacity of each other limit. (b) Compute the usable capacity gained by adding 3 MW of CRAH capacity, by adding 3 MW of airflow capacity, and by adding 5 MW of airflow capacity, and explain the last result with equation (17.17). (c) Under N−1 the electrical limit falls to 50 MW and the chilled-water plant to 45 MW. Compute the secure capacity and say which limit binds.

Exercise 17.2. A site has a 60 MW utility limit and cooling power $P_{\text{cool}} = 0.25\,P_{\text{IT}}$. (a) Compute the maximum IT load. (b) After the loss of one transformer the secure facility limit is 54 MW; compute the contingency-secure IT load. (c) On a hot afternoon the cooling power rises to $0.30\,P_{\text{IT}}$; repeat (a) and (b). (d) A generator plant of 55 MW firm capacity must carry the facility at its worst-case ratio of (c); compute its capacity in IT terms and say whether it is the bottleneck.

Table 17.4: The objects of capacity and risk and where each is defined.

Object	Expression	Role	Defined in
Usable capacity	$\min_i C_i$, equation (17.1)	set by the bottleneck	17.1.1
Capacity margin	$M = C - L$	headroom at stated conditions	17.1.2
State dependence	equations (17.3), (17.4)	capacity moves with weather and equipment	17.2.1
Secure capacity	equation (17.5)	feasible in every required contingency	17.2.4
Rack margin	equation (17.6)	local headroom, neighbor-dependent	17.3.2
Thermal capacity	equation (17.7)	capacity as an optimization	17.3.3
Coupled limit	$P_{\text{lim}}/1.2$, equation (17.10)	cooling power shares the electrical limit	17.3.6
Chance constraint	equation (17.11)	limit stated as a probability	17.4.2
Risk-constrained capacity	C_ϵ, equation (17.13)	largest load at tolerance ϵ	17.4.2
Normal margin risk	equation (17.14)	$\Phi(-\mathbb{E}[M]/\sigma_M)$	17.4.5
Quantile capacity	C_q, equation (17.15)	capacity exceeded with probability $q/100$	17.4.6
Forecast margin	$M(t+h)$	margin at a horizon	17.5.3
Expected risk	$\mathcal{R} = \Pr(M < 0)\,\ell$	probability weighted by consequence	17.5.6
Marginal usable capacity	equation (17.17)	value of an investment	17.6.3
Feasible region	$C = f(d, \theta, \text{state})$	capacity as a fuzzy surface	17.6.5

Exercise 17.3. The cooling capacity of a hall is $C \sim \mathcal{N}(6.0, 0.4^2)\,\mathrm{MW}$ and its heat load $L \sim \mathcal{N}(5.2, 0.3^2)\,\mathrm{MW}$, independent. (a) Compute $\mathbb{E}[M]$, σ_M and $\Pr(M < 0)$. (b) Compute the largest mean load, with the same σ_L, for which $\Pr(M < 0) \le 1\,\%$, using $z_{0.99} = 2.326$. (c) A model of the hottest inlet gives $\mu_T = 26.0 + 1.5\,(P_{\mathrm{IT}} - 5)^{\circ}\mathrm{C}$ with $\sigma_T = 0.5\,\mathrm{K}$ and $T_{\max} = 27^{\circ}\mathrm{C}$; compute the violation probability at $P_{\mathrm{IT}} = 5\,\mathrm{MW}$ and the $1\,\%$ capacity C_{99}.

Exercise 17.4. At 10:00 a facility has $C = 8.0\,\mathrm{MW}$ of cooling and $L = 6.0\,\mathrm{MW}$ of heat load. The forecast for 13:00 is $C = 6.9\,\mathrm{MW}$ with $\sigma_C = 0.3\,\mathrm{MW}$ and $L = 6.5\,\mathrm{MW}$ with $\sigma_L = 0.4\,\mathrm{MW}$. (a) Compute the present margin, the forecast mean margin and $\Pr(M < 0)$ at 13:00 for independent errors. (b) Repeat with a correlation $\rho = -0.5$ between the errors, and explain the sign of ρ. (c) A violation in this zone costs $\ell = \$400\,000$; compute the expected risk $\mathcal{R}$ for (b). (d) How much load must be shifted out of the 13:00 hour to bring $\Pr(M < 0)$ below $1\,\%$ in case (a), if shifting load reduces the mean of L but not its standard deviation?

17.9 Connection to the TERRA framework

Capacity is where the model of the facility built in the earlier chapters is used to answer a planning question. In the framework, every capacity in this chapter is a computation on the graph rather than a number read from a nameplate.

- **Capacities are component bounds on three coupled networks.** Transformers, UPS modules and PDUs are components of the electrical network; chillers, pumps and coils of the water and thermal networks; air handlers, containment and tiles of the air network. Each carries a bound. The usable capacity of equation (17.1) is the largest IT injection at the rack nodes for which the coupled solve keeps every component within its bound, and the coupling of Section 17.3.6, heat at a rack node becoming electrical load at the chiller, is the solve itself rather than a factor of 1.2.

- **Contingencies are topology changes.** A contingency c removes a component or an edge from the graph G; the secure capacity of equation (17.5) and the contingency-secure IT load of equation (17.8) are the same capacity computed on each modified graph, with the minimum taken over the contingency set.

- **Inferred parameters carry their uncertainty into capacity.** Rack airflow, the recirculation fraction α, the coil conductance UA and sensor biases b are inferred, with a posterior. Propagating samples of that posterior through the forward solve gives $p(C)$, and the capacity risk is the probability that available capacity is below required capacity, $\Pr(\dot{Q}_{\mathrm{avail}} < \dot{Q}_{\mathrm{req}}) = \Pr(M < 0)$, from which the quantile capacities C_{50}, C_{95}, C_{99} are read.

- **Forecasts are disturbance inputs in time.** Weather and load forecasts with their uncertainty, including the negative correlation of Section 17.5.2, drive a forward simulation; the result is $\Pr(M(t+h) < 0)$ for each horizon, as in Figure 17.7, and a slowly drifting posterior on UA is the degradation of Section 17.5.4 seen months before an alarm.

- **Attribution and the next measurement come from the same model.** The active bound at the capacity limit is the bottleneck of Table 17.2; perturbing a bound and re-solving gives the marginal usable capacity; perturbing a load injection gives the what-if chain of Figure 17.8; and the inferred parameter that contributes most to the spread of $p(C)$ names the measurement that the report of Table 17.3 recommends.

In summary, the capacity of a data center is the smallest of several coupled limits, each dependent on weather, equipment state and contingencies and each known only through

measured, calculated and inferred quantities; it is properly stated as a margin with a probability, $\Pr(M < 0)$, a risk weighted by consequence, and a quantile capacity at a chosen tolerance, recomputed as conditions and evidence change. The next chapter develops what-if analysis (Chapter 18): propagating a proposed change through the facility model to see which limits it reaches and with what probability.

Chapter 18

What-If Analysis

18.1 The what-if question

The preceding chapters asked two questions of a data center that is already operating. State estimation (Chapter 14) asks what is happening now: which temperatures, flows and loads the facility has at this moment, including the ones no sensor reads. Diagnosis (Chapters 13 and 16) asks why: which cause best explains a symptom. Operation also needs a third question, asked before anything has happened. The question of this chapter is

what would happen to the facility if something changed: a load, a setpoint, a piece of equipment, the weather, or the trust placed in a sensor?

The change is hypothetical. It may be a plan (adding a megawatt of IT load, raising a setpoint), a threat (a chiller trips on the hottest afternoon of the year), or a doubt (a sensor reads high and may be biased). In each case the answer is obtained the same way: change the inputs, controls, parameters or component states of a model of the facility, propagate the consequences through the coupled electrical, air, water and thermal systems, and compare the result with what the model predicts without the change. The chapter works through load, failure, weather and measurement scenarios, then turns to how uncertainty changes the form of the answer, how repeated what-if analysis becomes a comparison of actions, and when a model can be trusted to answer a given what-if question at all.

Jargon: what-if analysis, scenario, baseline, counterfactual, contingency, intervention, sensitivity analysis, licensing of a model

What-if analysis is the prediction of a system's behavior under a hypothetical change, made by running a model with the change and comparing the result with a run without it. A *scenario* S is one such hypothetical change, stated precisely enough to be put into the model. The *baseline* is the model's prediction with no change, the reference against which every scenario is compared. A *counterfactual* is a statement about what would happen, or would have happened, under conditions that do not hold; a what-if prediction is a counterfactual about the future. A *contingency* c is the loss of one or more components; an $N{-}1$ contingency is the loss of any single component, the term used in the redundancy levels of Chapter 4 and in the contingency analysis of a transmission system. An *intervention* is a change that an operator imposes on the system, as opposed to one that is merely observed to occur. *Sensitivity analysis* measures how much a model's output changes when one of its inputs or parameters changes. *Licensing* a model for a question means establishing, by tests specific to that question, that the model's answer can be relied on.

Key idea

A what-if analysis runs the same model twice, once as it is and once with a hypothetical change, from the same estimated present state, and reports the difference: in temperatures, powers and margins, in the time remaining before a limit, and, when the inputs are uncertain, in the probability of violating a limit. It is the bridge between understanding a facility and deciding what to do in it.

18.1.1 Three questions about one system

The three operational questions differ in their direction in time and in the object they produce. Write y for the record of measurements up to the present time t, x_t for the state of the facility at that time (temperatures, flows, stored heat, loads), H_i for a hypothesis about a cause, with H_0 normal operation, and u, d and θ for the control inputs, disturbances and parameters of a model. Then

- state estimation produces the distribution of the present state given the record, $p(x_t \mid y)$;
- diagnosis produces the posterior probability of each cause given the record, $\Pr(H_i \mid y)$;
- what-if analysis produces the distribution of a future state x_{t+h}, a horizon h ahead, under a change that has not happened.

The generic form of the what-if model is

$$x_{t+h} = f(x_t,\, u,\, d,\, \theta), \tag{18.1}$$

where f is the facility model of the earlier chapters stepped forward over the horizon: the electrical distribution of Chapter 3, the heat balance of Chapter 5, the airflow of Chapter 6, the cooling plant of Chapter 7, the storage and lags of Chapter 8 and the controls of Chapter 10. A scenario changes one or more arguments of f, and the analysis propagates the change through it.

Figure 18.1 places the three questions on one time axis. The horizontal direction is time, with the present marked “now”. The jagged line to the left of the present is the measured record y. To the right, the gray dashed line is the baseline prediction and the red line is the prediction under a scenario; the dashed red horizontal line is a limit. The three boxes above name the question asked of each region: diagnosis looks back along the record, estimation looks at the present, and what-if analysis looks forward along two branches that start from the same estimated state. Only the first two use data about the events they describe; the third uses data about the present and a model for everything after it, which is why the credibility of the model for the question becomes a topic of its own in Section 18.6.

18.1.2 Kinds of what-if question

What-if questions arise in planning, in operation and in the interpretation of data. Typical examples in a data center are: the IT load rises by 2 MW; chiller 3 fails; the outdoor wet-bulb temperature is 5 K higher than today; one UPS path is unavailable; the chilled-water supply temperature is raised; rack 12 is replaced by a 100 kW rack of AI accelerators; a temperature sensor is biased; one computer room air handler (CRAH) is lost; the airflow through row 7 drops by 20 %. Each is a counterfactual operating question: the facility is not in that condition, and the analysis asks what it would be like if it were.

The examples differ in what they change in equation (18.1), and that difference decides how the scenario is entered into a model. Table 18.1 sorts them. The first column is the question; the second names the argument of the model it changes; the third says whether the consequence is mainly facility-wide or local to a hall, row or rack. A load change is a change of input; a chiller or CRAH loss is a change of component state; a wet-bulb rise is a change of disturbance; a setpoint is a control; an airflow loss or a fouled coil is a parameter; a biased sensor changes not

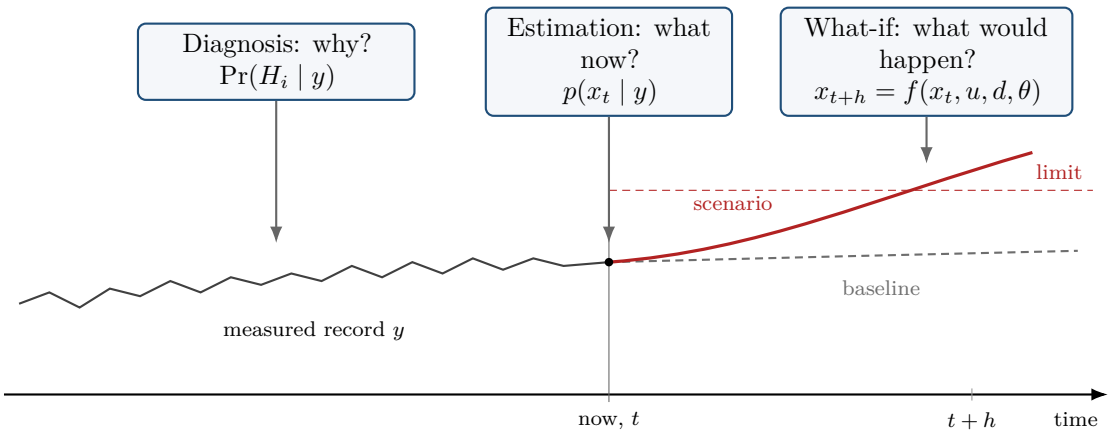


Figure 18.1: Three questions asked of one facility. Diagnosis explains the measured record, estimation fixes the present state, and what-if analysis predicts two futures from that state, the baseline and a scenario, and compares them (schematic).

the facility but its measurement. The same model answers all of them, provided it represents the argument that the question changes.

Table 18.1: Typical what-if questions in a data center, the argument of the model each changes, and the reach of its consequence.

What if	Changes	Reach
IT load rises by 2 MW	input: IT power P_{IT}	facility-wide
Rack 12 becomes a 100 kW AI rack	input: one rack's P_{rack}	local, rack
Chiller 3 fails	component state (contingency c)	facility-wide
One CRAH is lost	component state (contingency c)	local, hall or row
One UPS path is unavailable	component state, topology G	electrical, facility-wide
Outdoor wet-bulb is 5 K higher	disturbance d: T_{wb}	facility-wide
Chilled-water supply temperature is raised	control u: setpoint T_{set}	facility-wide
Airflow through row 7 drops by 20 %	parameter θ: flow resistance	local, row
A temperature sensor is biased	measurement: bias b	interpretation

18.1.3 Baseline and scenario

A what-if analysis needs a reference. The *baseline* x_0 is the model's description of the facility as it is, or as it is predicted to be over the horizon with no change. For the illustrative facility used throughout this chapter the baseline is

$$P_{\mathrm{IT}} = 8\,\mathrm{MW}, \qquad \dot{Q}_{\mathrm{avail}} = 11\,\mathrm{MW}, \qquad T_{\mathrm{rack,max}} = 26^{\circ}\mathrm{C}, \qquad P_{\mathrm{facility}} = 10\,\mathrm{MW}, \tag{18.2}$$

where $\dot{Q}_{\mathrm{avail}}$ is the cooling capacity available at the baseline's outdoor conditions and $T_{\mathrm{rack,max}}$ is the highest rack inlet temperature in the facility. The four numbers are of different kinds. P_{IT} and P_{facility} are measured, by meters on the IT feeds and on the utility service; the IT heat $\dot{Q}_{\mathrm{IT}} = P_{\mathrm{IT}} = 8\,\mathrm{MW}$ is calculated from the measured power by the energy balance of Chapter 5; the available capacity $\dot{Q}_{\mathrm{avail}}$ is calculated from a plant model and the present weather,

or inferred from the plant's recent behavior; the hottest rack temperature is measured if a sensor sits at the hottest rack and inferred otherwise. The baseline power usage effectiveness is PUE $= 10/8 = 1.25$, and the cooling margin is $\dot{Q}_{\text{avail}} - \dot{Q}_{\text{IT}} = 11 - 8 = 3\,\text{MW}$.

A scenario S_1 changes something and gives a predicted state x_1. The answer to the what-if question is the difference

$$\Delta x = x_1 - x_0, \tag{18.3}$$

taken component by component: the change in facility power, in cooling margin, in the hottest rack temperature, in the time remaining before a limit. Two runs are compared rather than one run read in isolation because the model's errors are then partly shared. A model that predicts the hottest rack 1 K too warm in the baseline tends to do so in the scenario as well, and the difference Δx is more reliable than either absolute prediction. The comparison is only fair if both runs start from the same estimated present state and use the same parameters except those the scenario changes; a scenario run from a different starting point mixes the effect of the change with the effect of the different start.

18.1.4 What a scenario may change

A scenario may change any of five things in the model, and a sixth outside it. It may change the inputs and controls u (an IT load, a setpoint, a fan speed command), the disturbances d (outdoor dry-bulb and wet-bulb temperature), the parameters θ (a coil's conductance UA, a filter's flow resistance, a recirculation fraction α), the topology G (which rack is fed by which power path, which CRAH serves which zone), and the availability of equipment, written as a contingency c, the set of components out of service. The sixth is the measurement model: a scenario in which a sensor is biased leaves the facility unchanged and changes what its readings mean.

Figure 18.2 shows where each kind of scenario enters the facility. The six boxes in the middle row are the stages through which a change propagates, in the order energy passes through them: IT load, electrical distribution, heat release into the air, the air and water systems that carry the heat, the plant and its controls, and the resulting risk to the service the IT equipment provides. The notes above each box list the scenarios of Table 18.1 that enter at that stage. The box below is the measurement system, which reads the air, water and electrical states and feeds the controls; a sensor bias enters there. The figure makes one property of what-if analysis visible: a change entered at the left propagates through every stage to its right, so a load scenario is also a cooling scenario and a service scenario, and a model that stops at the electrical stage answers only part of it.

18.2 Load scenarios

The simplest what-if questions change the IT load. This section adds one megawatt to the facility as a whole, then the same megawatt to a single hall, then a single high-density rack, and finally examines more airflow as the remedy for a hot rack, with its cost.

18.2.1 One more megawatt of IT load

Scenario S_1 raises the IT load from 8 to 9 MW. Because the electrical power of the IT equipment is converted to heat as it is consumed (Chapter 5), the IT heat rises with it, $\dot{Q}_{\text{IT}}$ from 8 to 9 MW. The added heat must be removed, so the chilled-water load rises, and with it the power of every machine in the cooling chain: the chillers, which must pump more heat up to the condenser; the chilled-water pumps, which must move more water at the same temperature difference; the CRAH fans, which must move more air at the same air temperature rise; and the cooling towers. The electrical distribution also loses a little more, since UPS and transformer

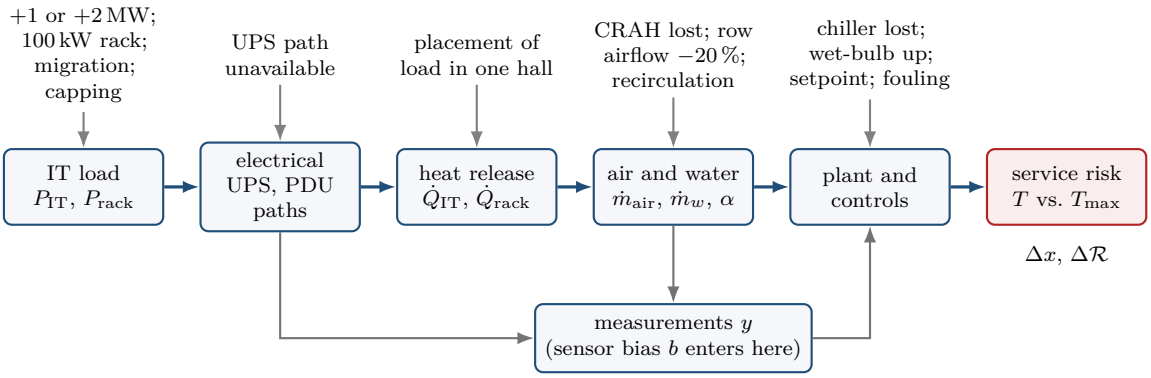


Figure 18.2: The propagation chain of a data center and where each kind of scenario enters it. A change entered at any stage propagates to every stage to its right; the measurement system reads the states and feeds the controls, so a sensor bias changes the control response without changing the physics (schematic).

losses grow with the power they carry. The change in facility power is the sum of all these increments.

Table 18.2 lists them for the illustrative facility. The first column is the component, the second the mechanism that sets its increment, and the third the increment in megawatts. The airflow and the water flow rise in proportion to the heat at fixed temperature differences, by $1/8 = 12.5\,\%$, and the power of a variable-speed fan or pump grows as the cube of its flow (the fan laws of Chapter 6), so each rises by a factor $1.125^3 = 1.42$. The chillers remove about 1.05 MW more heat (the IT megawatt plus the extra distribution, fan and pump heat) at an illustrative coefficient of performance of 8.3, and draw $1.05/8.3 \approx 0.13$ MW more. The total is

$$\Delta P_{\text{facility}} = 1.18\,\text{MW} \quad \text{for} \quad \Delta P_{\text{IT}} = 1\,\text{MW}, \tag{18.4}$$

so the facility power rises from 10 to 11.18 MW. The incremental ratio, 1.18, is lower than the average PUE of 1.25, because part of the overhead (lighting, offices, transformer no-load losses, the standby equipment) does not grow with the load. An incremental question needs the incremental ratio; multiplying the added load by the average PUE would overstate the added power by 0.07 MW.

Table 18.2: Increment of facility power when the IT load rises from 8 to 9 MW (illustrative). Fan and pump powers scale as the cube of flow; the chillers' increment is the extra heat divided by an illustrative COP = 8.3.

Component	Mechanism	ΔP (MW)
IT equipment	the scenario itself	1.000
UPS and distribution	losses of about 3 % of the added power	0.030
CRAH fans	airflow +12.5 %, power ×1.42 from 0.030 MW	0.013
Chilled-water pumps	water flow +12.5 %, power ×1.42 from 0.012 MW	0.005
Chillers	1.05 MW of extra heat at COP ≈ 8.3	0.126
Towers and condenser pumps	more heat rejected outdoors	0.006
Facility	sum	1.180

The scenario also moves the thermal state. With the airflow and water flow tracking the heat, the facility-level temperatures change little, and the hottest rack inlet rises by an illustrative

1 K, from 26 to 27°C, through slightly larger recirculation at the warmest racks. The cooling margin falls from 3 to $11 - 9 = 2$ MW. None of these numbers is measured before the load arrives: the increments of Table 18.2 are calculated from component models, and the change of the hottest rack temperature is a model prediction. The question "what if the IT load rises by 1 MW" is therefore a question about the coupled system, not about the IT equipment alone.

18.2.2 Where the load is placed

The same megawatt can be placed in different ways. Scenario S_2 adds it entirely to hall A, which already carries 2 MW. At the facility level nothing changes from S_1: the plant still has 11 MW of available cooling against 9 MW of heat, and $\dot{Q}_{\text{cool,total}} > \dot{Q}_{\text{req,total}}$. In hall A the situation is different. Its CRAHs can deliver at most an illustrative 200 kg/s of air. At the design temperature rise of 12 K across the racks, the racks of a 3 MW hall draw

$$\dot{m}_{\text{air}} = \frac{\dot{Q}}{c_{p,\text{air}}\,\Delta T} = \frac{3000\,\text{kW}}{1.0\,\text{kJ/(kg K)} \times 12\,\text{K}} = 250\,\text{kg/s},$$

which is 50 kg/s more than the CRAHs supply. The servers' fans draw the air they need regardless, and the difference comes from wherever it can: hot exhaust air pulled back over or around the racks into the cold aisle. Hall A's average recirculation fraction becomes $\alpha = 50/250 = 0.2$.

By the recirculation mixing rule of Chapter 6, a rack whose inlet air is a fraction α of its own exhaust and $1 - \alpha$ of supply air has an inlet temperature $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha(T_{\text{in}} + \Delta T)$, so

$$T_{\text{in}} = T_{\text{supply}} + \frac{\alpha}{1-\alpha}\,\Delta T. \tag{18.5}$$

With $T_{\text{supply}} = 20$°C and $\Delta T = 12$ K, the hall average is $20 + 0.25 \times 12 = 23$°C. The average hides the spread: racks at the ends of rows, where exhaust wraps around most easily, see an illustrative $\alpha = 0.4$, and their inlets reach $20 + (0.4/0.6) \times 12 = 28$°C, above an illustrative recommended inlet limit of 27°C. Hall A's hottest rack can thus exceed its limit while the facility has 2 MW of spare cooling. Where load is added matters as much as how much is added, and a facility-level model cannot see the difference between S_1 and S_2.

> **Principle 18.1.** *A facility-level cooling margin is necessary but not sufficient: the same added load can be safe when spread and unsafe when concentrated, because the local airflow, not the plant capacity, limits the heat that one hall, row or rack can reject.*

18.2.3 A 100 kW rack

Scenario S_3 replaces one 30 kW rack by a 100 kW rack of AI accelerators, an addition of 70 kW, less than 1 % of the plant's 11 MW. At the facility level it is negligible. At the rack it is not. With the rack's airflow unchanged at $\dot{m} = 3$ kg/s and $c_{p,\text{air}} \approx 1.0$ kJ/(kg K), the air temperature rise across the rack,

$$\Delta T = \frac{\dot{Q}_{\text{rack}}}{\dot{m}\,c_{p,\text{air}}}, \tag{18.6}$$

goes from

$$\frac{30\,\text{kW}}{3\,\text{kg/s} \times 1.0\,\text{kJ/(kg K)}} = 10\,\text{K} \quad \text{before to} \quad \frac{100\,\text{kW}}{3\,\text{kg/s} \times 1.0\,\text{kJ/(kg K)}} \approx 33\,\text{K} \quad \text{after.}$$

With a 20°C inlet the exhaust rises from 30 to about 53°C, well above what server components, cabling and the people who work in a hot aisle tolerate, and the hot exhaust also raises the

recirculation temperature seen by neighboring racks. Holding the rise at 12 K would need $100/(1.0 \times 12) = 8.3$ kg/s of air, about 7 m^3/s at 1.2 kg/m^3, nearly three times the original flow through the same rack footprint. The problem is local heat removal, not total chiller capacity: the plant has the capacity, and the air path to this rack does not. This is why racks of this density are usually cooled by liquid brought to the rack or to the chip rather than by room air.

18.2.4 More airflow and its fan power

Scenario S_4 asks what happens if the airflow through the 100 kW rack is doubled, from 3 to 6 kg/s. The temperature rise halves, to $100/6 \approx 16.7$ K, but the fan power grows as the cube of the speed, $P_{\text{fan}} \propto N_{\text{fan}}^3$, while the airflow grows only in proportion to it, $\dot{V} \propto N_{\text{fan}}$ (Chapter 6). Doubling the flow multiplies the fan power by $2^3 = 8$. If the rack's server fans draw an illustrative 1.0 kW at 3 kg/s, they draw 8 kW at 6 kg/s, an increase of 7 kW. That power is itself converted to heat inside the rack, so the rack now rejects about 107 kW and the rise is $107/6 = 17.8$ K rather than 16.7 K. The pressure drop across the rack also rises, as the square of the flow, and so does the tendency of air to leak around rather than through the equipment.

Figure 18.3 draws both sides of the trade. The left panel has the rack airflow on the horizontal axis and the air temperature rise across the rack on the vertical axis, for the original 30 kW rack and the 100 kW replacement; the dots mark the three operating points of this section. The right panel has the same horizontal axis and the server fan power on the vertical axis. The temperature rise falls as the inverse of the flow, so the first added kilogram per second buys the most; the fan power rises as the cube, so the last added kilogram per second costs the most. Doubling the flow moves the 100 kW rack from 33 to about 17 K of rise at eight times the fan power: a benefit and a cost, both of which a what-if analysis must report.

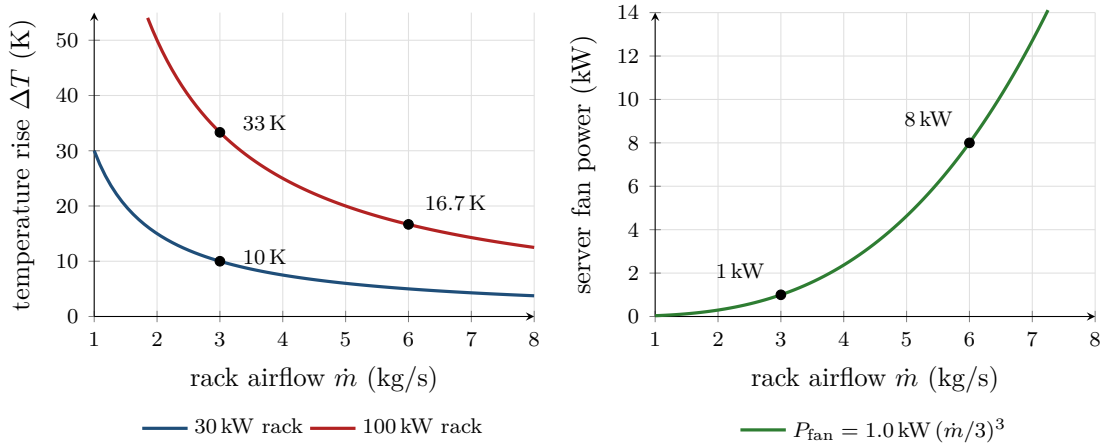


Figure 18.3: Left: air temperature rise across a rack, $\Delta T = \dot{Q}_{\text{rack}}/(\dot{m}\, c_{p,\text{air}})$ with $c_{p,\text{air}} = 1.0$ kJ/(kg K), for a 30 kW and a 100 kW rack. Right: server fan power under the cube law, $P_{\text{fan}} = 1.0\,\text{kW}\,(\dot{m}/3\,\text{kg/s})^3$ (illustrative). Doubling the flow of the 100 kW rack from 3 to 6 kg/s halves the rise and multiplies the fan power by eight.

18.3 Failure and weather scenarios

The most consequential what-if questions concern equipment that stops and weather that turns against the plant. This section takes the loss of one chiller through its consequence in time, its recovery and the probability that recovery fails, then the effect of hotter weather on both the power and the capacity of the plant, and finally the combination of the two.

18.3.1 Loss of a chiller

Take the facility after scenario S_1, with 9 MW of IT heat, and a plant of three chillers of 4 MW each, 12 MW at their rating. For round numbers this section takes the cooling requirement as the IT heat, $\dot{Q}_{\text{req}} = 9\,\text{MW}$, and neglects the few percent of distribution and fan heat of Table 18.2. Scenario S_5 is the loss of chiller 3. The available capacity drops to $\dot{Q}_{\text{avail}} = 2 \times 4 = 8\,\text{MW}$ and the plant falls 1 MW short:

$$\dot{Q}_{\text{req}} - \dot{Q}_{\text{avail}} = 9 - 8 = 1\,\text{MW}.$$

In steady state the scenario has no solution within limits: no temperature at which the plant removes 9 MW exists, because the two remaining chillers are at their maximum. A steady-state analysis can only report that the N−1 check fails.

The temperatures, however, do not jump. The heat that the plant cannot remove is stored in the thermal mass of the facility (the air, the racks, the building structure and the water in the loops), and by the dynamic heat balance of Chapter 8,

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\text{req}} - \dot{Q}_{\text{avail}}, \tag{18.7}$$

the shortfall sets the rate at which temperature rises, not its value. The operational question is therefore not whether the facility can sustain the loss, which it cannot, but how quickly the temperatures approach their limits and whether something can be done before they arrive.

18.3.2 Time to limit

Suppose the model, from equation (18.7) with the facility's effective thermal capacitance, predicts that the hottest rack inlet rises at 0.25°C/min after the loss. The implied capacitance is

$$C_{\text{th}} = \frac{\dot{Q}_{\text{req}} - \dot{Q}_{\text{avail}}}{\mathrm{d}T/\mathrm{d}t} = \frac{10^6\,\text{W}}{0.25/60\,\text{K/s}} = 2.4 \times 10^8\,\text{J/K} = 240\,\text{MJ/K}.$$

The hottest rack is at 27°C (the state after S_1) and its allowable limit is $T_{\max} = 35°\text{C}$. With the rise linear, because the remaining chillers are saturated and their heat removal no longer grows as the room warms, the time to limit is

$$t_{\text{limit}} = \frac{T_{\max} - T_0}{\mathrm{d}T/\mathrm{d}t} = \frac{35 - 27}{0.25}\,\text{min} = 32\,\text{min}. \tag{18.8}$$

Chapter 8 showed that the linear extrapolation is conservative when part of the heat removal still grows with temperature, and that the hottest rack can rise faster than the bulk when its local airflow is marginal; both refinements change the number and neither changes its form.

The statement "one chiller failure leaves about 32 minutes before the first thermal limit" is more useful than the statement "the facility fails the N−1 test". The first says how urgent the response is and which responses are fast enough; the second says only that some response is needed. A what-if analysis of a failure should produce a time, not a verdict.

18.3.3 Recovery by a standby chiller

The plant has a fourth chiller on standby that takes 8 min to start and load. If it starts, the available capacity returns to 12 MW and the race between recovery and the limit is won with $32 - 8 = 24\,\text{min}$ to spare. The hottest rack reaches $27 + 0.25 \times 8 = 29°\text{C}$ at the moment of recovery, after which the plant has $12 - 9 = 3\,\text{MW}$ of surplus and the temperature falls at $3 \times 0.25 = 0.75°\text{C/min}$ until the controls bring it back to its normal value.

The standby chiller does not always start. Starting is the moment of highest stress for a machine that has been idle, and a standby unit can fail to start for reasons ranging from a

failed start relay to a valve left closed after maintenance (the failure-on-demand probabilities of Chapter 12). With an illustrative probability $p_{\text{fs}} = 0.2$ that the standby fails to start, the outcome is no longer certain, and the what-if question becomes probabilistic: the probability that recovery takes longer than the time to limit,

$$\Pr(t_{\text{rec}} > t_{\text{limit}}) = p_{\text{fs}} + (1 - p_{\text{fs}})\ \Pr(t_{\text{start}} > t_{\text{limit}} \mid \text{start succeeds}). \tag{18.9}$$

The second term is the chance that a successful start is too slow. With a start time of 8 min against a 32 min limit it is negligible, and $\Pr(t_{\text{rec}} > t_{\text{limit}}) \approx 0.2$: if the standby chiller were the only response, one chiller trip in five would end with a rack above its limit. The reliability of the standby equipment and the thermal dynamics of the facility enter the same number, which neither a reliability study nor a thermal study gives alone. Section 18.4.6 adds a second line of defense and carries the calculation to the probability of each outcome.

Figure 18.4 draws the trajectories. The horizontal axis is minutes since chiller 3 tripped and the vertical axis is the hottest rack inlet temperature; the dashed red line is the limit $T_{\text{max}} = 35°\text{C}$. The solid red line is the loss with no response, reaching the limit at 32 min. The blue line is recovery by the standby chiller at 8 min. The green line is the case in which the standby fails and 1.2 MW of flexible IT load is shed instead, with the shedding complete at 18 min (Section 18.4.6). The dashed gray line is the combined scenario of Section 18.3.6, the same chiller loss on an extreme-weather day, with no response. All four start at the same temperature, and the first three coincide until a response takes effect; what separates them is the shortfall after the trip and the time at which it is closed.

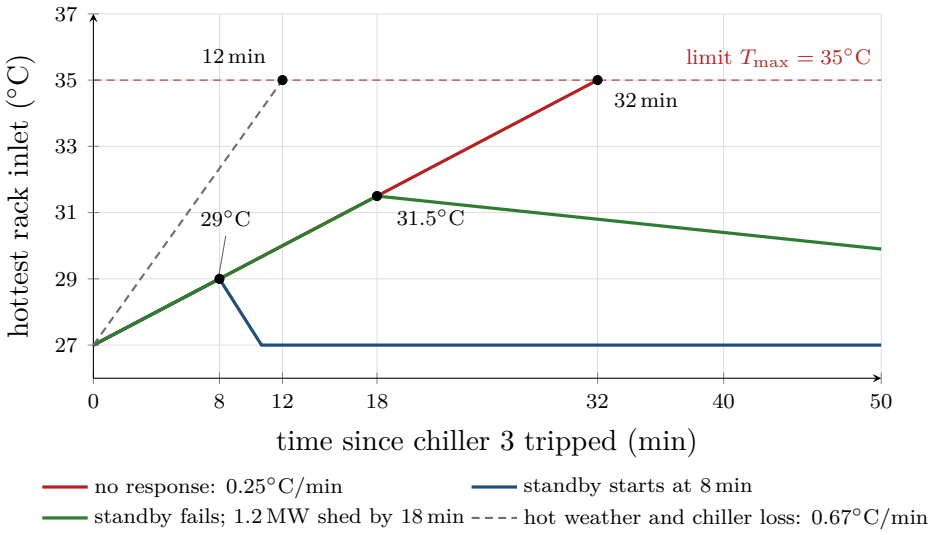


Figure 18.4: Hottest rack inlet after the loss of chiller 3 with $\dot{Q}_{\text{req}} = 9$ MW and $C_{\text{th}} = 240$ MJ/K, so that each megawatt of shortfall raises the temperature by 0.25°C/min (illustrative). Recovery by the standby chiller turns a 1 MW shortfall into a 3 MW surplus at 8 min; shedding 1.2 MW turns it into a 0.2 MW surplus at 18 min; the combined hot-weather case has a 2.67 MW shortfall.

18.3.4 Hotter weather

A rise in the outdoor wet-bulb temperature acts on the plant through a chain of dependencies (Chapter 7). Cooling towers cool the condenser water by evaporation, and evaporation can bring the water no lower than the wet-bulb temperature; a tower leaves its water a roughly constant approach above T_{wb}. A higher wet-bulb therefore raises the condenser-water supply temperature $T_{\text{CW,s}}$, which raises the chillers' condensing temperature T_{cond} and with it the lift, the temperature difference $T_{\text{cond}} - T_{\text{evap}}$ the compressor must pump heat across. A larger lift

lowers the coefficient of performance, so the same heat costs more compressor power; and since a compressor has a maximum power, it also lowers the maximum heat a chiller can remove. In words,

$$T_{\mathrm{wb}} \uparrow \;\rightarrow\; \text{tower performance} \downarrow \;\rightarrow\; T_{\mathrm{CW,s}} \uparrow \;\rightarrow\; \text{lift} \uparrow \;\rightarrow\; \text{COP} \downarrow \text{ and } \dot{Q}_{\mathrm{chiller,max}} \downarrow .$$

An illustrative calculation makes the chain concrete. Model the chiller's COP as a fixed fraction $\eta = 0.75$ of the ideal (Carnot) value for its two temperatures, $\mathrm{COP} = \eta\, T_{\mathrm{evap}}/(T_{\mathrm{cond}} - T_{\mathrm{evap}})$ with temperatures in kelvin. At the reference conditions the evaporator runs at 5°C (278.15 K) and the condenser at 30°C, a lift of 25 K, so $\mathrm{COP} = 0.75 \times 278.15/25 = 8.3$. A wet-bulb 5 K higher, with the tower approach and the condenser approach unchanged, raises the condensing temperature to 35°C and the lift to 30 K, so $\mathrm{COP} = 0.75 \times 278.15/30 = 6.95$, a drop of 17 %. Two consequences follow together. At the 9 MW load the chillers' power rises from $9/8.3 = 1.08$ MW to $9/6.95 = 1.29$ MW, by 0.21 MW. And if each chiller is limited by its compressor's maximum power, its maximum heat removal falls in the same proportion as its COP, and the plant's 12 MW becomes $12 \times 6.95/8.3 \approx 10.0$ MW. Hotter weather means more cooling power and less cooling capacity at the same time, which is the least favorable combination for a facility near its limits.

18.3.5 A weather scenario and its margin

Scenario S_6 is an extreme-weather day, slightly more severe than the 5 K example, for which the plant model gives a maximum capacity of $\dot{Q}_{\mathrm{cool,max}} = 9.8$ MW against 12 MW at the reference conditions. The cooling requirement on that day is 9.2 MW: the 9 MW of IT heat plus an illustrative 0.2 MW of heat that hot weather adds, through the building envelope and through fans and pumps that work harder. The margin falls from

$$12 - 9.2 = 2.8\,\mathrm{MW} \quad \text{to} \quad 9.8 - 9.2 = 0.6\,\mathrm{MW}.$$

The facility remains within capacity, and a steady-state check passes. But its tolerance of a contingency has gone: on that day each chiller can deliver $9.8/3 \approx 3.3$ MW, and the loss of any one of them removes more than five times the remaining margin. A facility that is N+1 on a mild day is N on an extreme day. The weather scenario does not produce a violation; it removes the protection against one.

18.3.6 Combined scenarios and nonlinear interaction

Scenarios occur together, and their combined consequence can exceed the sum of their separate ones. Take the hot day of S_6 together with the chiller loss of S_5. The two remaining chillers deliver $2 \times 9.8/3 = 6.53$ MW against 9.2 MW, a shortfall of 2.67 MW. With the capacitance of Section 18.3.2 the hottest rack rises at

$$\frac{2.67 \times 10^{6}\,\mathrm{W}}{2.4 \times 10^{8}\,\mathrm{J/K}} = 0.0111\,\mathrm{K/s} = 0.67^{\circ}\mathrm{C/min},$$

and reaches 35°C in $8/0.67 = 12$ min, the dashed line of Figure 18.4. Separately, the chiller loss gave a 1 MW shortfall and 32 min, and the hot day gave no shortfall at all. The standby chiller still starts in time, at 8 min with the hottest rack at $27 + 0.67 \times 8 = 32.3$°C, and its derated 3.3 MW turns the shortfall into a 0.6 MW surplus; but the time to spare has fallen from 24 to 4 min, and a standby failure now leaves 4 min, not 24, for a second response. The number of scenario combinations grows quickly with the number of components and conditions, and choosing which combinations to evaluate is part of the analysis (Section 18.4.6).

Interaction. A combination is more than a sum when one change alters the effect of another. A fouled condenser is an example. Fouling adds a thermal resistance on the water side of the condenser, so the condensing temperature must rise further above the condenser water to reject the same heat; the higher the condenser-water temperature already is, the closer the chiller already runs to its lift and pressure limits, and the more the added rise costs. Fouling therefore matters more on a hot day. Model it illustratively: weather alone removes 2 % of capacity per kelvin of wet-bulb rise, fouling alone removes 10 % at the reference conditions, and the fouling penalty grows by 15 % of itself per kelvin of wet-bulb rise. At a 5 K rise, weather alone removes 10 %, fouling at that wet-bulb removes $10 \times (1 + 0.15 \times 5) = 17.5$ %, and the combined loss is

$$1 - (1 - 0.10)(1 - 0.175) = 1 - 0.9 \times 0.825 = 0.2575,$$

about 26 %, against 20 % for the sum of the separate effects and 19 % for their product.

Figure 18.5 draws the illustrative model. The horizontal axis is the rise of the wet-bulb temperature above the reference and the vertical axis is the loss of chiller capacity in percent. The blue line is weather alone with a clean condenser; the dotted gray line is what an additive analysis would predict for weather plus fouling, the weather loss plus a constant 10 %; the red line is the combined loss with the interaction. The red and gray lines start together at no weather change and separate as the weather worsens; at 5 K the gap is about six percentage points, enough to turn a 0.6 MW margin into a shortfall in a plant of this size.

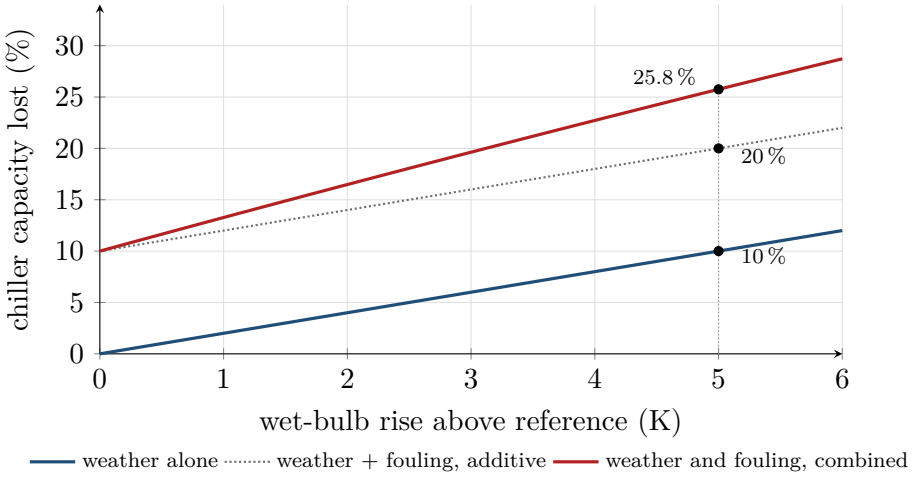


Figure 18.5: Interaction of hot weather and condenser fouling (illustrative). Weather alone: 2 % per kelvin of wet-bulb rise ΔT_{wb}. Fouling: $10\,\%\,(1 + 0.15\,\Delta T_{\mathrm{wb}})$. Combined: $1 - (1 - 0.02\,\Delta T_{\mathrm{wb}})(1 - 0.10(1 + 0.15\,\Delta T_{\mathrm{wb}}))$. The additive line assumes the fouling penalty does not depend on the weather.

> **Principle 18.2.** *Combined scenarios are not sums of individual scenarios: a change that removes a margin turns a second, survivable change into a violation, and a degradation whose effect grows with operating conditions makes the combined loss larger than the losses added.*

18.4 Uncertainty in scenarios

Every number in the previous sections was a single prediction. This section admits that the state, the parameters and even the measurements are uncertain: it treats a doubt about a sensor as a scenario, uses what-if runs to test hypotheses, and shows how a deterministic prediction becomes a probability, how the probability of a scenario differs from its consequence, and how branching events form a scenario tree.

18.4.1 A sensor bias as a scenario

A scenario need not change the facility; it can change what the facility's readings mean. Rack 7's inlet sensor reports 31°C. A reading is the true value plus a bias plus noise, $y = x + b + v$, with v the measurement noise, and two scenarios account for the reading equally well:

- scenario A, $b = 0$: the inlet is truly at 31°C;
- scenario B, $b = +4$ K: the sensor reads 4 K high, and the inlet is at 27°C.

The two scenarios lead to different thermal states, different risks and different control responses. Under A the rack is 4 K from an allowable limit of 35°C; a chiller loss at 0.25°C/min would leave it 16 min, half the 32 of Section 18.3.2, and the right control response is to raise local airflow or lower the supply temperature now. Under B the rack is 8 K from the limit, the cooling is adequate, and the right response is to schedule a check of the sensor; raising the airflow would spend fan power to cool a rack that is not hot. A controller that acts on the reading as if it were the truth takes the action of scenario A whichever is true.

The useful form of the question is not "is the sensor biased?", which the reading alone cannot answer, but "if this sensor were biased, what else should be observed?". A what-if run under each scenario predicts the readings of every other sensor, and those predictions can be compared with what the other sensors show.

18.4.2 What-if as hypothesis checking

Simulate scenario B, $b = +4$ K, and ask whether the other readings become mutually consistent: the neighboring racks' inlet temperatures, the rack's exhaust temperature, the heat balance of the rack, the airflow, and the inlet temperatures reported by the servers' own management controllers. If they do, the bias hypothesis gains credibility; if they fit scenario A instead, the reading is probably true.

Table 18.3 carries out the check for an illustrative 20 kW rack with 1.67 kg/s of air, so a 12 K rise. The first column is the quantity, the second says how it is obtained, the next two give its predicted value under each scenario, and the last gives the value observed. The rack-7 inlet sensor itself cannot discriminate: both scenarios reproduce its 31°C by construction. Every other row discriminates. Under A the exhaust should be $31 + 12 = 43$°C, the servers' own inlet sensors should read about 31°C, the server fans should have sped up, and the neighbors, which share the same air path, should be warm too; under B all of these should look normal. The observations agree with B in every row. A single biased sensor explains everything; scenario A needs the server sensors, the exhaust sensor and the neighbors' sensors all to be wrong at once.

Table 18.3: Predicted readings under two scenarios for rack 7 (inlet sensor reads 31°C; illustrative 20 kW rack with a 12 K rise), compared with the observed values.

Quantity	Kind	A: $b = 0$	B: $b = +4$ K	Observed
Rack 7 inlet sensor	measured	31°C	31°C	31°C
Rack 7 server inlet sensors	measured	about 31°C	about 27°C	27.4°C
Rack 7 exhaust	measured	43°C	39°C	39.3°C
Rack 7 server fan speed	measured	raised	normal	normal
Racks 6 and 8 inlets	measured	about 29°C	about 26.5°C	26.7°C
Rack 7 heat balance	calculated	closes	closes	–
True rack 7 inlet	inferred	31°C	27°C	–

The heat balance row shows a limit of the check: the rack's power and its exhaust-minus-inlet difference cannot discriminate if the inlet reading is the one in doubt, since each scenario supplies its own inlet. Discrimination comes from independent sensors whose predictions differ. Turning

the comparison into a posterior probability, $\Pr(H_B \mid y)$ against $\Pr(H_A \mid y)$, with the noise of each sensor and the prior rate of biased sensors, is the subject of Chapter 16; the what-if run supplies the predictions that the posterior compares.

18.4.3 Raising the chilled-water setpoint

Scenario S_7 raises the chilled-water supply setpoint from 7 to 10°C. The evaporator temperature rises with it, the lift falls by about 3 K, and with the Carnot-fraction model of Section 18.3.4 the chillers' COP rises from $0.75 \times 278.15/25 = 8.3$ to $0.75 \times 281.15/22 = 9.6$, about 15 %. Warmer chilled water also lengthens the hours in which a waterside economizer can carry part of the load, since the condenser water is more often colder than the return chilled water. Against these savings, the CRAH coils deliver warmer supply air or need more airflow for the same supply temperature, so the racks warm and the fans work harder.

For an illustrative facility whose cooling system draws 2.0 MW, a plant and airflow model predicts

$$P_{\text{cool}} : 2.0 \;\rightarrow\; 1.7\,\text{MW}, \qquad T_{\text{rack,max}} : 27 \;\rightarrow\; 29^{\circ}\text{C}, \qquad T_{\max} = 35^{\circ}\text{C}.$$

The change saves 0.3 MW, 15 % of the cooling power, and leaves the hottest rack 6 K below its allowable limit. As a deterministic prediction the change is attractive, and an analysis that stopped here would recommend it.

18.4.4 Deterministic and probabilistic statements

The prediction $T_{\text{rack,max}} = 29^{\circ}\text{C}$ is uncertain. The recirculation fractions behind it are inferred, the coil conductances drift, the rack powers fluctuate, and the model has errors of its own. Suppose the combined uncertainty is a standard deviation of $\sigma_T = 3\,\text{K}$, roughly normal. The right question is then not whether $T_{\text{rack,max}} < 35^{\circ}\text{C}$, which the point prediction answers yes, but how probable a violation is:

$$\Pr(T_{\text{rack,max}} > T_{\max}) = 1 - \Phi\!\left(\frac{T_{\max} - \mu_T}{\sigma_T}\right) = 1 - \Phi\!\left(\frac{35 - 29}{3}\right) = 1 - \Phi(2) = 0.023, \quad (18.10)$$

where $\mu_T = 29^{\circ}\text{C}$ is the predicted mean and Φ is the standard normal distribution function. A 2.3 % probability of exceeding the limit is a different statement from "6 K of margin". Whether it is acceptable depends on a risk tolerance ϵ: under $\epsilon = 0.05$ the change is acceptable, under $\epsilon = 0.01$ it is not. The tolerance also fixes how much uncertainty the decision can bear. For $\epsilon = 0.01$ the margin of 6 K must be at least $2.33\,\sigma_T$, since $1 - \Phi(2.33) = 0.01$, so the change is acceptable only if $\sigma_T \le 6/2.33 = 2.6\,\text{K}$. Reducing the uncertainty, for example by measuring the recirculation at the hottest racks, can turn a rejected change into an accepted one without changing the facility. This is a risk-aware what-if analysis: the scenario is judged by a probability of violation, not by a predicted value.

The same distinction applies to failure scenarios. A deterministic statement reads "if chiller 3 fails, the maximum rack temperature reaches 33°C". A probabilistic one reads "if chiller 3 fails, the maximum rack temperature has a median of 33°C and a 9 % probability of exceeding 35°C". The second carries more information: under a normal distribution it implies a spread of $\sigma_T = (35 - 33)/\Phi^{-1}(0.91) = 2/1.34 = 1.5\,\text{K}$, and it tells an operator that roughly one chiller loss in eleven ends above the limit, which the first statement hides entirely.

Figure 18.6 draws the two distributions. The horizontal axis is the predicted maximum rack temperature and the vertical axis is probability density. The blue curve is the setpoint scenario, centered at 29°C with $\sigma_T = 3\,\text{K}$; the red curve is the chiller-loss scenario, centered at 33°C with $\sigma_T = 1.5\,\text{K}$. The dashed vertical line is the limit, and the shaded areas to its right are the probabilities of violation. The setpoint scenario has the lower mean and the wider spread, the chiller scenario the higher mean and the narrower spread; both have tails beyond the limit, and the tail, not the mean, is what the decision depends on.

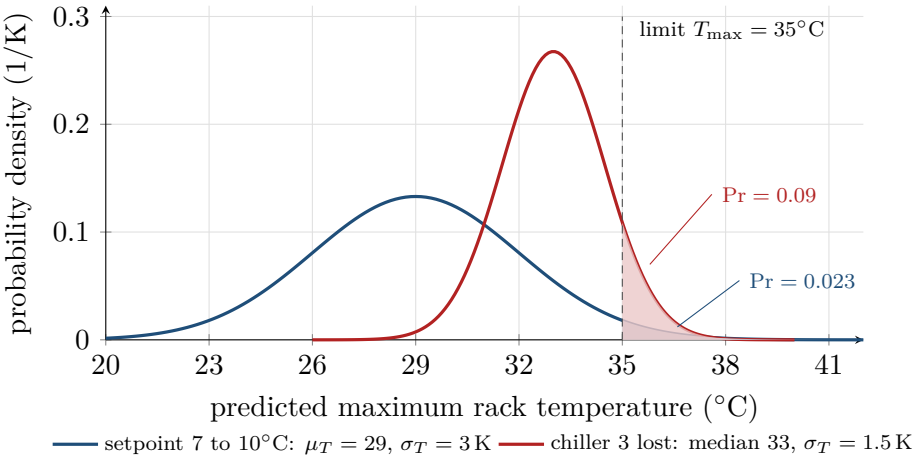


Figure 18.6: Predicted maximum rack temperature as a distribution for two scenarios (illustrative, normal): the setpoint change, $\mathcal{N}(29, 3^2)$, with $\Pr(T > 35°\mathrm{C}) = 1 - \Phi(2) = 0.023$; and the chiller loss, $\mathcal{N}(33, 1.5^2)$, with $\Pr(T > 35°\mathrm{C}) = 0.09$. Shaded: the probability of exceeding the limit.

18.4.5 Scenario probability and consequence

Two probabilities must be kept apart. The first is the distribution of the consequence given that the scenario occurs, $p(\ell \mid S)$, where ℓ is a loss: degraded service, equipment damage, energy cost. What-if analysis computes it; the tail probabilities of Section 18.4.4 are summaries of it. The second is the probability that the scenario occurs at all, $\Pr(S)$, which comes from reliability data and weather statistics; risk analysis supplies it (Chapters 12 and 17). The expected risk over a set of scenarios combines the two,

$$\mathcal{R} = \sum_{S} \Pr(S)\, \mathbb{E}[\ell \mid S], \tag{18.11}$$

the familiar statement that risk is probability times consequence. A what-if analysis alone says how bad a scenario is if it happens, and nothing about how often it happens.

Two scenarios of different kinds. The simultaneous loss of both chilled-water pumps is severe: water stops moving, the plant delivers no cooling to the halls, and the time to limit is a few minutes. Its illustrative probability is $\Pr(S) = 10^{-6}$ per year, because the pumps are redundant and independently powered. A 10 % degradation of airflow through a row, from a clogged filter or a misplaced floor tile, has a moderate consequence, a few kelvin at a few racks and some fan energy, but happens several times a year. With illustrative consequences of 5×10^6 cost units for the pump loss and 2×10^4 for an airflow degradation occurring with probability 0.3 per year, the expected risks are 5 and 6000 cost units per year. By expected risk the airflow degradation is a thousand times more important. The two are nevertheless evaluated differently: the frequent, moderate scenario is managed by monitoring and maintenance, where expected cost is the right measure; the rare, severe one is managed by design and by standing procedures, because its consequence is unacceptable whatever its probability, and an operator holds its probability below a tolerance rather than trading it against energy savings.

18.4.6 Scenario trees

A failure rarely ends with the first event. Each response can succeed or fail, and each outcome leads to further branches. A scenario tree lays out the sequence: the chiller fails; the standby starts or fails to start; if it fails, flexible IT load is shed successfully or the shedding fails. Each

branch has a probability, a temperature trajectory $T(t)$, and a consequence for the service the IT equipment provides. The probability of a leaf is the product of the probabilities along its path.

Figure 18.7 draws the tree for the chiller loss of Section 18.3.3. The box on the left is the initiating event; the edges are labeled with the conditional probability of each branch; the boxes on the right are the outcomes, with the peak temperature of their trajectory (from Figure 18.4), their consequence and their probability. With illustrative probabilities 0.8 that the standby starts and 0.9 that load shedding succeeds when it is needed, the three outcomes have probabilities

$$0.8, \qquad 0.2 \times 0.9 = 0.18, \qquad 0.2 \times 0.1 = 0.02.$$

Shedding 1.2 MW is chosen so that the plant's 8 MW exceeds the remaining 7.8 MW of heat: once the shedding completes at 18 min, with the rack at $27 + 0.25 \times 18 = 31.5$°C, a 0.2 MW surplus cools it at 0.05°C/min. The second line of defense reduces the probability of reaching the limit from 0.2, with the standby chiller alone, to 0.02. The leaves sum to one, which is a useful check that no branch was dropped.

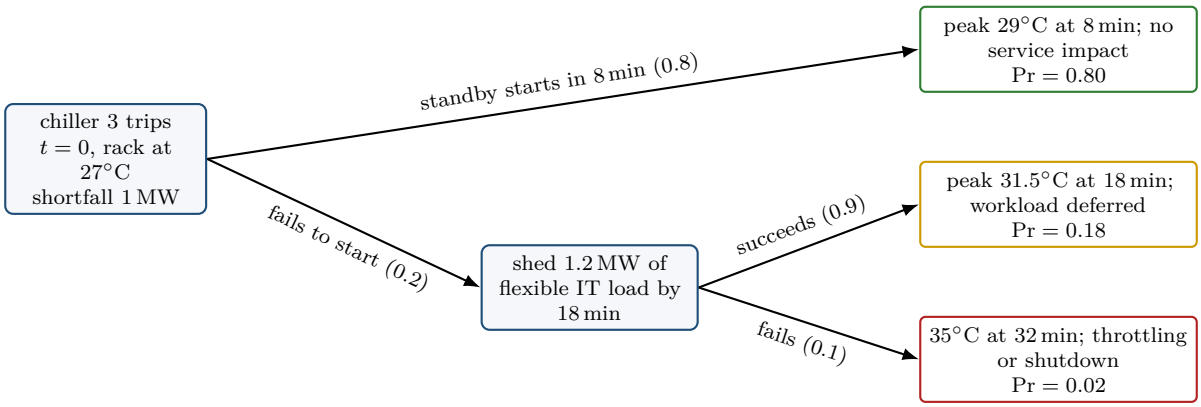


Figure 18.7: Scenario tree for the loss of chiller 3 (illustrative probabilities). Each leaf's probability is the product of the conditional probabilities along its path; its peak temperature comes from the trajectories of Figure 18.4.

A tree turns the what-if analysis of single scenarios into a risk statement for the initiating event: given that chiller 3 trips, the probability of a thermal limit violation is 0.02, and the probability of some service impact is 0.2. Multiplying by the probability that a chiller trips at all, from Chapter 12, gives the contribution of this event to the facility's annual risk. Trees grow quickly, and in practice branches whose probability falls below a threshold are pruned; the threshold is part of the analysis and should be reported with it.

18.5 Actions and sensitivity

A what-if analysis can change the operator's own actions as well as the facility's conditions. This section examines two actions, moving workload and capping power, compares actions as a small optimization, separates an intervention from a forecast, and then asks which inputs a prediction is sensitive to and in what order constraints become active as the load grows.

18.5.1 Workload migration

Return to the placement scenario S_2, in which hall A carries 3 MW and its end-of-row racks reach 28°C. Scenario S_8 moves 1 MW of flexible compute, batch jobs or training runs that can run anywhere, from hall A to hall B. Hall A's power falls, $\Delta P_A = -1$ MW, and hall B's rises,

$\Delta P_B = +1\,\mathrm{MW}$; the total IT load is unchanged. Hall A returns to 2 MW, within its 200 kg/s of CRAH air, its recirculation falls toward zero, and its hottest inlet returns toward the supply temperature. Hall B warms: if it carried 3 MW with CRAHs able to deliver 300 kg/s, its racks now draw $4000/(1.0 \times 12) = 333\,\mathrm{kg/s}$, and it is hall B's turn to recirculate. The migration helps only if the receiving hall has the airflow and cooling to spare.

The total cooling power changes too, though the total heat does not. The CRAH fans of hall A slow down and those of hall B speed up, and since fan power is cubic in flow, moving airflow from a hall running near its maximum to one running below it saves fan power: illustratively, two halls at 90 % and 60 % of their maximum airflow draw $0.9^3 + 0.6^3 = 0.95$ units of fan power, and the same total airflow split as 75 % and 75 % draws $2 \times 0.75^3 = 0.84$. A migration is a what-if scenario with a thermal, an electrical and an efficiency consequence, and all three should be predicted before the jobs move.

18.5.2 Power capping

Scenario S_9 caps the power of 20 GPU racks from 100 to 80 kW each. The electrical demand falls by

$$\Delta P = 20 \times (80 - 100)\,\mathrm{kW} = -400\,\mathrm{kW},$$

and since the IT power becomes heat, the heat to be removed falls by about 400 kW as well (a little more, counting distribution and fan losses). During the chiller outage of Section 18.3.1 the cap reduces the 1 MW shortfall to 0.6 MW, the rate of rise from 0.25 to $0.6 \times 0.25 = 0.15$°C/min, and the time to limit from 32 to $8/0.15 = 53$ min. On a day when the shortfall is smaller than 0.4 MW, the cap removes it entirely and prevents the violation.

The cost is computation. A GPU running at a capped power delivers less throughput, though usually less than in proportion to the power cut, because the last watts of a processor buy the least performance; an illustrative 20 % power cap costs about 10 % of throughput on the capped racks. A power cap is a control action whose consequences cross from the facility into the service, and the what-if analysis must report both sides.

18.5.3 Comparing actions

After the loss of chiller 3, an operator has several candidate actions a:

- a_1: push the remaining chillers above their rating, by an illustrative 8 %, adding 0.6 MW of capacity;
- a_2: power-cap the 20 GPU racks to 80 kW, removing 0.4 MW of heat;
- a_3: migrate 0.8 MW of flexible workload to another site, which takes 15 min to complete.

Each is simulated forward, x_{t+h} under a_i, and the predictions are compared on thermal risk, electrical demand, cost and computing performance. Table 18.4 gives the comparison. Each row is an action; the columns give the shortfall that remains, the resulting rate of rise and time to limit from 27°C to 35°C with $C_{\mathrm{th}} = 240\,\mathrm{MJ/K}$, the change in electrical demand, and the cost on the service side. For a_3 the first 15 min are spent at the full rate, reaching 30.75°C, and the remaining 4.25 K at the reduced rate, so $t_{\mathrm{limit}} = 15 + 4.25/0.05 = 100$ min.

Choosing among the actions is an optimization. Write $\ell(a)$ for a loss that combines the criteria with weights set by the operator's policy, for example $\ell(a) = w_1 \Pr(T > T_{\max} \mid a) + w_2\,\Delta P(a) + w_3\,\mathrm{cost}(a) + w_4\,\Delta\mathrm{performance}(a)$, and choose

$$a^\star = \arg\min_{a_i} \; \mathbb{E}\big[\ell(a_i)\big], \tag{18.12}$$

the expectation taken over the uncertain state and parameters. No row of Table 18.4 dominates: a_1 keeps all computing running but adds demand and risks a second trip; a_2 acts at once but costs throughput; a_3 gives the longest time to limit but takes a quarter of an hour. Which is best depends on the weights, and the weights are a policy, not a result of the model. The structural

Table 18.4: Candidate actions after the loss of chiller 3 (illustrative; $\dot{Q}_{\text{req}} = 9\,\text{MW}$, $\dot{Q}_{\text{avail}} = 8\,\text{MW}$, hottest rack at 27°C, limit 35°C).

Action	Shortfall (MW)	Rate (°C/min)	t_{limit} (min)	Demand (MW)	Service and equipment cost
None	1.0	0.25	32	0	none until the limit
a_1: chillers +8 %	0.4	0.10	80	+0.15	wear; risk of a second trip
a_2: cap 20 GPU racks	0.6	0.15	53	−0.40	about 10 % throughput on 20 racks
a_3: migrate 0.8 MW	0.2	0.05	100	−0.80	jobs moved; transfer delay
a_2 and a_3	−0.2	falls	none	−1.20	both of the above

point is that repeated what-if analysis over a set of candidate actions is an optimization problem, solved here by enumeration; the systematic version is the subject of Chapter 20.

18.5.4 Forecast and intervention

A what-if question about an action is different from a forecast. A forecast predicts what is likely to happen given what has been observed, $p(x_{t+h} \mid y)$. A what-if analysis of an action predicts what would happen if the operator set a control to a chosen value,

$$p(x_{t+h} \mid y,\ \text{do}(u = u')), \tag{18.13}$$

where $\text{do}(u = u')$ denotes an intervention: the control is set to u' by the operator, rather than observed to take that value. The distinction matters because, in an operating facility, controls are chosen by controllers in response to the state, so observing a control value carries information about the state. Historical data from a facility show that high fan speeds accompany high rack temperatures, because the fan controllers speed up when racks are hot. Conditioning on observing a high fan speed therefore predicts a hot rack; setting the fan speed high, an intervention, makes the rack cooler. The do-notation marks this difference: an intervention cuts the link from the state to the controller and keeps only the physical effect of the control on the state. A model of physics, in which the fan speed enters the heat balance through the airflow, answers the interventional question directly; a model fitted to the correlations in operating data answers the observational one, and used for a what-if question it can predict the wrong sign.

18.5.5 Sensitivity

A sensitivity analysis asks how much a prediction changes when one input changes, with the others held fixed. Take the hottest rack inlet as a function of the recirculation fraction, from equation (18.5) with an illustrative supply temperature $T_{\text{supply}} = 20°\text{C}$ and a rack rise of $\Delta T = 15\,\text{K}$:

$$T_{\text{rack,max}}(\alpha) = T_{\text{supply}} + \frac{\alpha}{1-\alpha}\,\Delta T, \qquad \frac{\partial T_{\text{rack,max}}}{\partial \alpha} = \frac{\Delta T}{(1-\alpha)^2}. \tag{18.14}$$

Varying α from 0 to 0.3 sweeps the inlet from 20 to $20 + (0.3/0.7) \times 15 = 26.4°\text{C}$. The derivative is the local sensitivity: at $\alpha = 0.1$ it is $15/0.81 = 18.5\,\text{K}$ per unit of α, or $0.185\,\text{K}$ per percentage point of recirculation; at $\alpha = 0.3$ it is $15/0.49 = 30.6\,\text{K}$ per unit, 65 % larger.

Figure 18.8 draws the relation. The horizontal axis is the recirculation fraction and the vertical axis is the inlet temperature; the dashed line is the tangent at $\alpha = 0.1$, whose slope

is the local sensitivity there. The curve bends upward: the sensitivity itself grows with α, so a rack that already recirculates is more sensitive to further recirculation than a clean one. A local sensitivity is a property of an operating point, and a derivative computed at one point does not describe the response far from it.

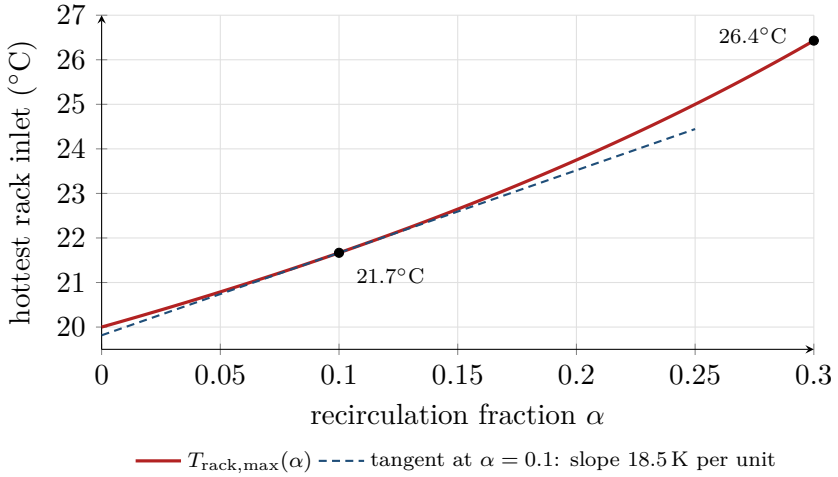


Figure 18.8: Hottest rack inlet as a function of the recirculation fraction, $T_{\mathrm{rack,max}} = T_{\mathrm{supply}} + \alpha\,\Delta T/(1-\alpha)$ with illustrative $T_{\mathrm{supply}} = 20°\mathrm{C}$ and $\Delta T = 15\,\mathrm{K}$, and the tangent at $\alpha = 0.1$, whose slope $\Delta T/(1-\alpha)^2 = 18.5\,\mathrm{K}$ is the local sensitivity.

18.5.6 Sensitivity and uncertainty

A large sensitivity does not make a parameter important to the uncertainty of a prediction. A highly sensitive parameter that is well known contributes little; a less sensitive one that is very uncertain may contribute much. What matters is the product of sensitivity and uncertainty. To first order, for independent parameters θ_j with standard deviations σ_j,

$$\mathrm{Var}(T_{\mathrm{rack,max}}) \approx \sum_j \left(\frac{\partial T_{\mathrm{rack,max}}}{\partial \theta_j} \right)^2 \sigma_j^2, \tag{18.15}$$

and the share of each term in the sum ranks the parameters by how much of the prediction's uncertainty they cause.

Table 18.5 applies equation (18.15) to the prediction of Figure 18.8 at $\alpha = 0.1$. Each row is a parameter: the supply air temperature, which is measured; the bias of the supply-air sensor, which is unknown; the recirculation fraction, which is inferred; and the rack temperature rise, which is calculated from power and estimated airflow. The columns give the sensitivity, the standard deviation, their product and the share of the variance. The supply temperature has the largest sensitivity of the three temperature-like parameters, 1 K per kelvin against 0.111 for the rack rise, but because it is measured to 0.3 K it contributes 7 %. The recirculation fraction has a sensitivity of only 0.185 K per percentage point, yet an uncertainty of 5 percentage points makes it 70 % of the variance. The total standard deviation is $\sqrt{1.23} = 1.1\,\mathrm{K}$, and the way to reduce it is to learn the recirculation, not to improve the supply sensor. This ranking is the starting point for the choice of the next measurement in Chapter 19.

18.5.7 The sequence of active constraints

A what-if analysis run repeatedly with a growing input finds the order in which the facility's constraints become active. For the facility of this chapter, raising the IT load step by step gives:

Table 18.5: First-order uncertainty budget of the hottest rack inlet at $\alpha = 0.1$, $T_{\text{supply}} = 20°\text{C}$, $\Delta T = 15\,\text{K}$ (illustrative standard deviations).

Parameter θ_j	Kind	$\partial T/\partial\theta_j$	σ_j	product (K)	share
Supply temperature T_{supply}	measured	$1\,\text{K}/\,\text{K}$	$0.3\,\text{K}$	0.30	7 %
Supply-sensor bias b	unknown	$1\,\text{K}/\,\text{K}$	$0.5\,\text{K}$	0.50	20 %
Recirculation α	inferred	$18.5\,\text{K}$ per unit	0.05	0.93	70 %
Rack rise ΔT	calculated	$0.111\,\text{K}/\,\text{K}$	$1.5\,\text{K}$	0.17	2 %
Total				1.11	100 %

+1 MW, every constraint satisfied; +2 MW, the airflow in hall C reaches its limit; +3 MW, the electrical capacity of the distribution reaches its limit. The useful output is not a single capacity number but the sequence of binding constraints, each with the load at which it binds. For expansion planning it says what to relieve first: adding CRAH capacity to hall C buys one megawatt before the electrical limit is reached, and relieving the electrical limit is pointless until hall C is relieved.

A larger illustrative facility with a nameplate of 100 MW, operating at $P_{\text{IT}} = 70\,\text{MW}$, shows the pattern in full. At 72 MW no constraint is active. At 75 MW the airflow margin of hall B becomes small. At 78 MW hall B reaches its thermal limit. At 82 MW the UPS system reaches its N−1 limit, the load beyond which the loss of one UPS module would overload the rest. At 85 MW the utility supply reaches its limit. Figure 18.9 draws the sequence. Each row is a constraint; the bar runs from the present load of 70 MW to the load at which the constraint becomes active; the dotted vertical line marks the 72 MW at which nothing binds. The facility's nameplate is 100 MW, its usable capacity without changes is 78 MW, and the gap between them is distributed over four specific constraints, each with its own remedy. "Nameplate 100 MW" says much less than this picture.

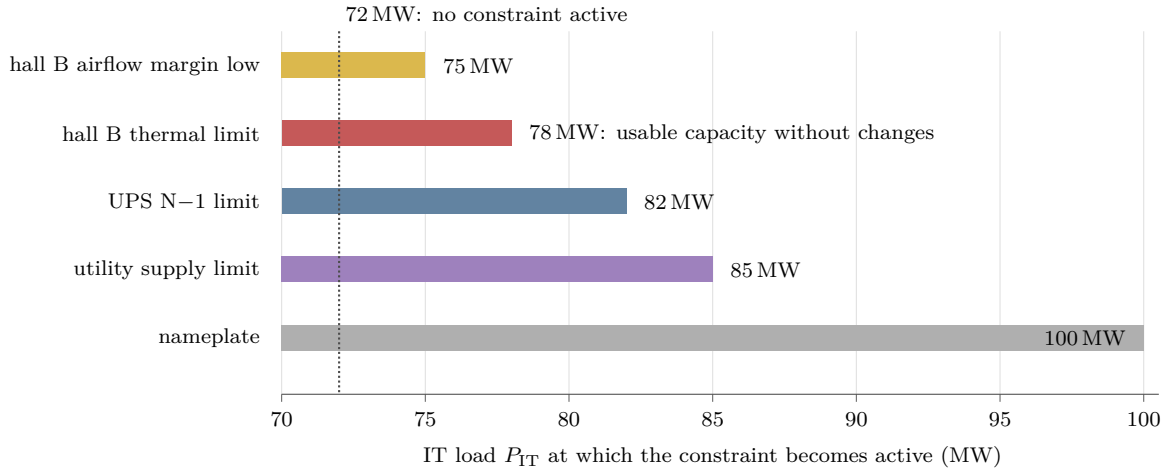


Figure 18.9: Constraint activation sequence for an illustrative facility of 100 MW nameplate operating at 70 MW. Each bar runs from the present load to the IT load at which its constraint becomes active. The first binding constraint, the thermal limit of hall B at 78 MW, sets the usable capacity.

Principle 18.3. *The capacity of a facility is a sequence, not a number: repeated what-if analysis with growing load finds the constraints in the order they bind, and the first one sets the usable capacity.*

18.6 Credibility of the model for the question

A what-if analysis predicts a condition that has not been observed, so its answer is only as credible as the model that produces it. This section describes what fidelity a scenario needs, why adequacy is a property of the question rather than of the model, how a model is licensed question by question, what a complete answer to a practical what-if question contains, and the mathematical structure common to all of them.

18.6.1 Model fidelity

A model that omits a mechanism cannot predict a scenario that turns on it, and the omission does not announce itself. Three omissions recur. A model without recirculation predicts that hall A of Section 18.2.2 is fine, because the facility has spare cooling; it is optimistic. A model without chiller derating predicts that the plant keeps 12 MW on the extreme day of Section 18.3.5; it is optimistic again, and it misses the loss of contingency tolerance. A model without transient thermal storage predicts that the chiller loss of Section 18.3.1 violates the limit at once, leaving no time for the standby; it is pessimistic and would recommend shedding load that the standby could have saved. In each case the scenario misleads, and in different directions.

Before a model is used for a class of what-if questions, three properties should be tested for that class. *Adequacy*: the model contains the mechanisms the scenario exercises (recirculation for placement scenarios, derating for weather scenarios, storage for failure scenarios). *Identifiability*: the parameters of those mechanisms can be determined from the data available, the subject of Chapter 15; a recirculation term whose coefficient the data cannot pin down adds a mechanism without adding credibility. *Validation*: the model has reproduced observed events of the same class, a past chiller trip, a hot day, a past load addition, within a stated error. A model that passes the three tests for one class has not passed them for another.

18.6.2 Adequacy for the question

Adequacy depends on the question, not only on the model. A simple thermal network, with one node per hall and lumped heat balances, is adequate for the question “what happens to the cooling margin if the total IT load rises by 10 %?”: the answer depends on totals, the plant model and the heat balance, all of which a hall-level network represents. The same network is useless for the question “which rack develops a hot spot if one perforated floor tile is moved?”: the answer depends on the pressure field under the floor and the jet of air from each tile, which a lumped node does not represent at all. That question may need computational fluid dynamics (CFD), a numerical solution of the equations of air motion on a fine three-dimensional grid, which Chapter 21 treats.

The same facility can thus need two models of very different resolution, and neither is better in general. A CFD model is too slow and has too many uncertain boundary conditions to run the thousands of scenarios of a risk analysis; a thermal network cannot resolve a tile. The choice is made per question, by asking which mechanisms the answer depends on and at what spatial and temporal scale.

18.6.3 Licensing by question

The outcome of the tests of Section 18.6.1, applied per question class, can be summarized as a license: a statement of which questions a given model is trusted to answer. For a facility model built as a calibrated thermal network with a plant model and hall-level airflow, Table 18.6 gives an illustrative license. The first column is the question class, the second the verdict, and the third the reason. Questions about totals and about the first minutes after a plant failure are

licensed, because the model contains the mechanisms and has been validated on events of that kind. Row-level temperature responses are licensed conditionally, where the row's recirculation has been inferred from data. Rack-scale turbulent recirculation is not licensed: the model has no mechanism for it.

Table 18.6: An illustrative license for a calibrated thermal-network facility model with a plant model and hall-level airflow.

Question class	Verdict	Reason
Total cooling margin under load or weather change	licensed	totals, plant derating and heat balance represented; validated on hot days
Approximate time to limit after a plant failure	licensed	storage represented; capacitance identified from past transients
Chiller-loss scenario	licensed	plant model validated on a recorded chiller trip
Row-level temperature response	conditionally	only for rows whose recirculation fraction is identified from data
Rack-scale turbulent recirculation, effect of one tile	not licensed	no spatial airflow mechanism; needs a CFD model

A license of this kind belongs with every what-if answer. An answer given for an unlicensed question is not necessarily wrong, but it has no evidence behind it, and it should be reported as such or not at all.

18.6.4 Repeated what-if under uncertainty

The questions an operator actually asks combine everything in this chapter. A typical one reads: if tomorrow's wet-bulb temperature is 3 K higher than forecast and chiller 2 is unavailable, what is the probability that hall C exceeds its thermal limit, and which action most reduces it? Answering it integrates a weather forecast (the distribution of tomorrow's wet-bulb, shifted by 3 K), the equipment state (chiller 2 out, entered as a contingency), the physical model (plant derating, hall C's airflow and recirculation, storage), the uncertainty in the present state and parameters (their posterior from estimation), a risk measure (the probability of exceeding the limit), and control (the candidate actions).

The computation is a loop of what-if runs. Draw many joint samples of the uncertain quantities, tomorrow's wet-bulb profile from the shifted forecast and the state and parameters from their posterior; for each sample, simulate the facility through the day with chiller 2 out; count the fraction of samples in which hall C exceeds its limit; that fraction estimates the probability. Then repeat the whole loop under each candidate action, with the same samples, so that the difference between actions is not hidden by the difference between sample sets, and report the action with the largest reduction in probability per unit of cost.

18.6.5 A practical output

The product of such an analysis is a short statement an operator can act on. Table 18.7 gives an illustrative example for the scenario "loss of chiller 2 at 16:00 during forecast peak conditions". The first column names each element of the answer; the second gives its value; the third says what kind of quantity it is. The cooling margin falls from +2.1 to −0.7 MW, a drop of 2.8 MW, which is the derated capacity of chiller 2 at peak conditions. The first rack limit is expected between 18 and 31 min after the loss, a range rather than a number because the capacitance and the starting temperatures are uncertain. The probability of exceeding 35°C within 30 min is 0.42. The best mitigation is to reduce hall B's flexible IT load by 0.8 MW, which turns the

-0.7 MW margin into $+0.1$ MW; the residual risk of 0.06 remains because a margin of 0.1 MW lies within the uncertainty of the capacity itself.

Table 18.7: An illustrative what-if output for an operator: loss of chiller 2 at 16:00 during forecast peak conditions.

Element	Value	Kind
Scenario	chiller 2 lost at 16:00, forecast peak	stated (input to the model)
Cooling margin	$+2.1 \rightarrow -0.7$ MW	calculated from inferred capacity
First rack limit	expected in 18 to 31 min	predicted, with uncertainty
$\Pr(T_{\max} > 35°\text{C}$ within 30 min)	0.42	inferred probability
Best mitigation	reduce hall B flexible IT load by 0.8 MW	chosen by comparing actions
Residual risk	0.06	inferred probability after action

Each line of the table rests on a different part of the chapter: the margin on the plant and weather scenarios, the time on the dynamic balance, the probability on the uncertainty of Section 18.4.4, the mitigation on the comparison of actions, and the whole on a model licensed for chiller-loss questions. A statement of this form is what distinguishes a decision aid from a simulation.

18.6.6 The mathematical structure

All the scenarios of the chapter share one structure. The baseline evolves as

$$x_{k+1} = f(x_k,\, u_k,\, d_k,\, \theta;\; G,\, c), \tag{18.16}$$

with k the time step, G the topology and c the contingency (here none). A scenario evolves as

$$x'_{k+1} = f(x'_k,\, u'_k,\, d'_k,\, \theta';\; G',\, c'), \qquad x'_0 = x_0, \tag{18.17}$$

with the same function f, the same initial state, and primed arguments wherever the scenario differs. The answer is the pair

$$\Delta x_k = x'_k - x_k, \qquad \Delta\mathcal{R} = \mathcal{R}' - \mathcal{R}, \tag{18.18}$$

the change in state along the horizon and the change in risk, where the risk $\mathcal{R}$ is computed from the distribution of the trajectory, for example as the probability of exceeding a limit within the horizon. A scenario may change the controls u, the disturbances d, the parameters θ, the topology G and the equipment availability c; a measurement scenario changes the map from x to y instead. Under uncertainty, x_0 and θ are drawn from their posterior $p(x_t, \theta \mid y)$, both runs use the same draws, and Δx and $\Delta\mathcal{R}$ become distributions. Every example of the chapter, from the added megawatt to the practical output of Table 18.7, is equation (18.17) with a particular choice of primed arguments, propagated along the chain of Figure 18.2.

18.7 What to retain

Definition 18.1 (What-if analysis)**.** What-if analysis predicts the consequence of a hypothetical change, a load, a failure, the weather, a setpoint, an action or a doubt about a sensor, by

running the same facility model with and without the change from the same estimated state and reporting the difference in states, in time to limit and in the probability of violating a limit; it bridges understanding the system and deciding what to do, by propagating the change through IT load, electrical distribution, heat, air and water, controls and service risk.

Table 18.8 lists the chapter's objects and the subsection that defines each.

Table 18.8: The objects of what-if analysis and where each is defined.

Object	Expression	Role	Defined in
What-if model	$x_{t+h} = f(x_t, u, d, \theta)$	predicts a future under a change	18.1.1
Baseline and difference	$\Delta x = x_1 - x_0$	the answer, with shared errors	18.1.3
Incremental facility power	$\Delta P_{\text{facility}} = 1.18\,\text{MW}$ per MW	coupled load scenario	18.2.1
Placement	equation (18.5)	local limit with global margin	18.2.2
Rack temperature rise	$\Delta T = \dot{Q}_{\text{rack}}/(\dot{m} c_{p,\text{air}})$	local heat removal	18.2.3
Fan cube law	$P_{\text{fan}} \propto N_{\text{fan}}^3$	cost of airflow	18.2.4
Shortfall dynamics	$C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = \dot{Q}_{\text{req}} - \dot{Q}_{\text{avail}}$	failure sets a rate	18.3.1
Time to limit	equation (18.8)	a time, not a verdict	18.3.2
Recovery probability	equation (18.9)	reliability with dynamics	18.3.3
Weather chain	$T_{\text{wb}} \uparrow \Rightarrow \text{COP} \downarrow,\ \dot{Q}_{\max} \downarrow$	more power, less capacity	18.3.4
Interaction	Figure 18.5	combined is not the sum	18.3.6
Bias scenario	$y = x + b + v$	hypothesis checking	18.4.2
Tail probability	equation (18.10)	risk-aware what-if	18.4.4
Expected risk	equation (18.11)	probability times consequence	18.4.5
Scenario tree	Figure 18.7	branch probabilities	18.4.6
Action comparison	equation (18.12)	what-if as optimization	18.5.3
Intervention	$p(x_{t+h} \mid y, \text{do}(u = u'))$	not a forecast	18.5.4
Sensitivity	equation (18.14)	local derivative	18.5.5
Uncertainty budget	equation (18.15)	sensitivity times uncertainty	18.5.6
Constraint sequence	Figure 18.9	capacity as a sequence	18.5.7
License	Table 18.6	trust per question class	18.6.3
Scenario structure	equations (18.16)–(18.18)	common form of all scenarios	18.6.6

18.8 Exercises

Exercise 18.1. A rack of 40 kW is replaced by one of 60 kW. Its airflow is 4 kg/s and $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$. (a) Compute the air temperature rise across the rack before and after. (b) Compute the airflow needed to hold the rise at 12 K after the replacement. (c) The rack's server fans draw 0.8 kW at 4 kg/s. Compute their power at the airflow of (b) by the cube law, and the temperature rise once the added fan power is counted as heat in the rack. (d) The facility has 3 MW of spare cooling. Explain why that number does not answer whether the replacement is safe.

Exercise 18.2. A plant has four chillers of 3 MW and a cooling requirement of 10.2 MW. The facility's effective thermal capacitance is 300 MJ/K, the hottest rack is at 28°C, and the limit is 35°C. (a) One chiller fails. Compute the shortfall, the rate of rise in °C/min and the time to limit. (b) A standby chiller takes 10 min to start and fails to start with probability 0.15. If it starts, what temperature does the hottest rack reach? Compute $\Pr(t_{\text{rec}} > t_{\text{limit}})$ when the standby is the only response. (c) If the standby fails, flexible load can be shed with success probability 0.85. Draw the scenario tree and compute the probability of each leaf. (d) On a hot day each chiller delivers only 2.6 MW. Repeat (a) and comment on why the combined consequence exceeds the sum of the separate ones.

Exercise 18.3. A setpoint change is predicted to bring the hottest rack to $\mu_T = 30$°C, with a normal uncertainty of $\sigma_T = 2.5$ K, against a limit of 35°C. (a) Compute $\Pr(T_{\text{rack,max}} > 35\text{°C})$. (b) With a risk tolerance $\epsilon = 0.01$, is the change acceptable? What is the largest σ_T for which it would be? (c) A chiller-loss study reports a median maximum of 32°C and a 5 % probability of exceeding 35°C. Assuming normality, compute the implied σ_T. (d) Write the deterministic and the probabilistic statement of the result of (c), and say what an operator learns from the second that the first hides.

Exercise 18.4. The hottest rack inlet is predicted by $T = T_{\text{supply}} + \alpha\,\Delta T/(1-\alpha)$ with $T_{\text{supply}} = 18$°C, $\Delta T = 12$ K and $\alpha = 0.15$. (a) Compute T and the sensitivities $\partial T/\partial T_{\text{supply}}$, $\partial T/\partial\alpha$ and $\partial T/\partial\Delta T$. (b) The standard deviations are 0.2 K for T_{supply}, 0.08 for α and 2 K for ΔT. Compute the first-order standard deviation of T and the variance share of each parameter. (c) Which single parameter would a new measurement most usefully pin down, and why is it not the one with the largest sensitivity per kelvin? (d) Operating records show that the supply temperature is lowest when racks are hottest. Explain why a model fitted to that correlation would mispredict the effect of lowering T_{supply}, using the distinction between $p(x \mid y)$ and $p(x \mid y, \text{do}(u = u'))$.

18.9 Connection to the TERRA framework

What-if analysis is forward simulation of the facility graph with some of its elements edited; every object of the chapter is an edit of a particular kind.

- **A scenario is an edit of the graph.** A load scenario changes a source on an electrical node; a setpoint changes a control on a component; a weather scenario changes a boundary disturbance on the heat-rejection node; a fouled coil changes a component parameter such as UA; a lost chiller or CRAH removes a component (the contingency c); a re-fed rack changes the topology G. Equation (18.17) is the same graph stepped forward with the edit, from the same estimated state as the baseline.

- **Propagation follows the coupled networks.** An added megawatt enters the electrical network, appears as a heat source on the rack nodes, is carried by mass flow on the air and water edges, and loads the plant components, so $\Delta P_{\text{facility}}$ of Table 18.2 is the sum of the changes on every component the three networks share. Placement is resolved only if halls and racks are nodes with their own airflow edges and recirculation edges.

- **Failure scenarios are simulations with storage and discrete component states.** The time to limit comes from thermal nodes with capacitance, as in Figure 18.4; the standby chiller and the load shedding are components whose in-service state switches with a probability, which turns one simulation into the scenario tree of Figure 18.7.

- **Uncertain scenarios are run from the posterior.** The present state, the recirculation fractions and the sensor biases are inferred with posteriors; drawing them and simulating

gives $\Pr(T > T_{\max})$ instead of a point prediction, and a bias scenario is a hypothesis whose predicted readings on the other sensors are compared with the measurements, as in Table 18.3.

- **Actions are interventions on control edges.** An action sets a control and cuts its controller, the do-operation of equation (18.13); comparing actions is repeated simulation under each, and the constraint sequence of Figure 18.9 is the order in which component bounds bind as a source grows.
- **The license is a property of the graph's resolution.** A graph with one node per hall answers margin and time-to-limit questions and is silent about tiles; the license of Table 18.6 records which questions the graph's nodes, edges and identified parameters support.

In summary, a what-if analysis runs the facility model twice from the same estimated state, once as it is and once with a hypothetical change in load, equipment, weather, setpoint, action or measurement, and reports the difference as a change in state, a time to limit and a probability of violation; the change propagates through the coupled electrical, air, water and thermal systems, combined changes interact, the answer under uncertainty is a probability, repeated what-if analysis over actions is an optimization, and the whole is credible only for questions the model is licensed to answer. The next chapter uses the uncertainty that remains in these predictions to choose the measurement that would reduce it most.

Chapter 19

Next-Best Measurement

19.1 An ambiguous diagnosis

Chapters 13 and 16 ended a diagnosis with a ranked list: a set of hypotheses H_i about the cause of a symptom, each with a posterior probability $\Pr(H_i \mid y)$ given the measurement record y. Often that list does not single out one cause. Two or three explanations remain plausible, the measurements in hand cannot tell them apart, and each explanation calls for a different response. This chapter's question is

when the current data cannot distinguish competing explanations, which measurement, test or safe intervention should be made next, and how is its value computed?

The answer turns diagnosis from a passive reading of telemetry into an active choice. A candidate measurement is valued before it is taken, by predicting what it would show under each hypothesis and asking how much those predictions differ, how much the resulting change of belief would alter the action that follows, and what the measurement costs. That is a value-of-information problem, and the chapter develops it from the information measure to the decision measure, then to tests that perturb the system, to sequences of measurements, and to the form in which a recommendation is reported.

Jargon: predictive distribution, entropy, bit, information gain, value of information, active diagnosis, dual control, experiment design, Fisher information

The *predictive distribution* of a candidate measurement is the probability distribution of the value it would return, computed before it is taken, under each hypothesis or averaged over the hypotheses. *Entropy* is a measure of how spread a probability distribution is; measured with base-2 logarithms its unit is the *bit*, the uncertainty of a fair choice between two equally likely alternatives. The *expected information gain* of a measurement is the reduction of entropy it is expected to produce, averaged over its possible outcomes. The *value of information* (VOI) is the reduction of expected loss, in money or another unit of consequence, that a measurement is expected to produce because it lets a better action be chosen. *Active diagnosis* is diagnosis that changes the system on purpose, for example by stepping a fan speed, so that the response separates the hypotheses. *Dual control* is control that serves two purposes at once: keeping the system performing and learning about it. *Optimal experiment design* is the choice of inputs and measurements that makes the data most informative about unknown parameters. *Fisher information* is a matrix that measures how sensitive the expected measurements are to the parameters, scaled by the measurement noise; its inverse bounds the precision with which the parameters can be estimated.

Key idea

A measurement is worth taking when the hypotheses still in play predict different values for it and when the result could change the action; its value is the expected reduction of uncertainty, or better, of decision loss, minus its cost. The next-best measurement is chosen, taken, used to update the posterior, and the choice is then made again.

19.1.1 An ambiguous overheating rack

Consider a rack whose inlet temperature sensor has risen about 4 K above its usual reading (an illustrative figure). The diagnostic reasoning of Chapter 16 has produced three hypotheses and their posterior probabilities:

$$\Pr(H_1 \mid y) = 0.45, \qquad \Pr(H_2 \mid y) = 0.40, \qquad \Pr(H_3 \mid y) = 0.15, \tag{19.1}$$

where H_1 is low airflow through the rack (failing server fans or an obstructed intake), H_2 is hot-air recirculation (exhaust air drawn back to the rack front over the top of the rack or around the row end), and H_3 is a bias of the inlet sensor itself (the rack is not hot; the reading is). The three probabilities sum to one because the three hypotheses are treated as exhaustive and mutually exclusive.

A probability of 0.45 is not a diagnosis. The most probable explanation is still more likely to be false than true: the odds against H_1 are 0.55 to 0.45. Acting on H_1 alone would send a technician to the server fans and leave a recirculation problem untouched four times in ten. The three hypotheses also call for different responses: H_1 calls for a fan replacement or the removal of an obstruction, H_2 for blanking panels or containment repairs, and H_3 for a recalibration and no thermal action at all. The diagnosis is ambiguous, and the productive question is not which hypothesis to bet on but which additional observation would separate them most.

19.1.2 Candidate measurements

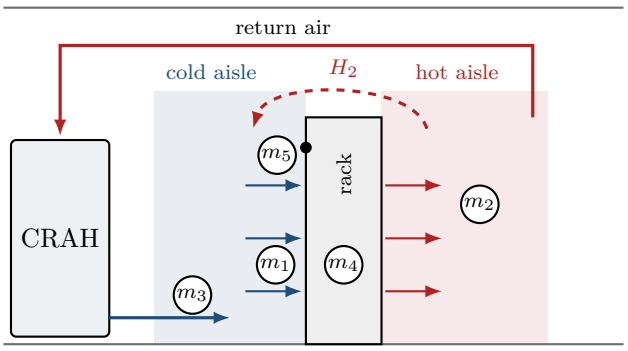
Five candidate measurements are available:

- m_1: a direct measurement of the rack airflow $\dot{m}_{\text{air}}$, for example with a flow hood or an anemometer traverse at the rack face;
- m_2: the hot-aisle air temperature behind the rack;
- m_3: the cold-aisle static pressure relative to the room, read from a differential pressure sensor at a floor tile;
- m_4: the speed of the server fans, reported by the servers' own management interface;
- m_5: a second, independent inlet temperature reading at the same rack, from a temporary calibrated probe.

Each has a cost: some are already in the telemetry, some need a technician, and one may need access inside the rack.

Figure 19.1 places them on a side view of the rack. The cold aisle is on the left, shaded blue, the rack in the middle and the hot aisle on the right, shaded red. Cooled air from the computer room air handler (CRAH) enters the cold aisle along the floor, passes through the rack from front to back, and returns over the ceiling. The dashed red arc over the rack top is the recirculation path of H_2. The black dot on the rack front is the inlet sensor that raised the alarm, whose bias is H_3. The circled labels mark where each candidate measurement is taken.

Table 19.1 lists, for each candidate, what it is predicted to show under each hypothesis and what it costs. The predictions follow from the physics of the earlier chapters. Low rack airflow



Hypotheses
H_1: low airflow through the rack
H_2: hot exhaust recirculated to the inlet (dashed arc)
H_3: bias of the alarming inlet sensor (black dot)

Candidate measurements
m_1: rack airflow $\dot{m}_{\text{air}}$ at the rack face
m_2: hot-aisle temperature
m_3: cold-aisle static pressure
m_4: server fan speed (telemetry)
m_5: second, independent inlet temperature

Figure 19.1: Side view of the ambiguous rack, schematic. Supply air enters the cold aisle (blue), passes through the rack and leaves into the hot aisle (red); the dashed arc is the recirculation path of H_2, the black dot the alarming inlet sensor of H_3. Circled labels mark the five candidate measurements.

lowers $\dot{m}_{\text{air}}$ directly. Recirculation leaves $\dot{m}_{\text{air}}$ unchanged but raises the inlet air temperature through the mixing rule $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{hot}}$ of Chapter 6. A biased sensor changes nothing physical, so every quantity except the biased reading itself looks normal. Server fans respond to the servers' own internal temperatures: they speed up when the rack really is hot (H_1 through an obstruction, H_2) and stay at their usual speed under H_3. The hot-aisle sensor sees the mixed exhaust of the whole row, so a change at one rack is diluted by its neighbors.

Table 19.1: Candidate measurements for the ambiguous rack: predicted outcome under each hypothesis and cost. Values and costs are illustrative.

Candidate	Under H_1	Under H_2	Under H_3	Cost and access
m_1 rack airflow	low, about 1.5 kg/s	normal, about 3.0 kg/s	normal, about 3.0 kg/s	technician with a flow hood; rack access, possibly a short shutdown
m_2 hot-aisle temperature	about 38°C	about 39°C	about 38.5°C	already measured; no cost
m_3 cold-aisle pressure	normal	slightly low	normal	portable manometer, minutes of technician time
m_4 server fan speed	high	high	normal	already in server telemetry; no cost
m_5 second inlet reading	hot	hot	normal	temporary calibrated probe, about 30 seconds on site

Reading the table by rows shows which pairs of hypotheses each candidate separates. m_1 separates H_1 from the other two but cannot separate H_2 from H_3, because both predict a normal airflow. m_5 and m_4 separate H_3 from the other two but cannot separate H_1 from H_2. m_2 separates nothing well, and m_3 separates weakly. No single candidate separates all three; the pair m_1 and m_5, taken together on one visit to the rack, does.

19.1.3 Predictive distributions of a measurement

The rows of Table 19.1 are point predictions. A measurement result is uncertain even when the hypothesis is known: the instrument has noise, and the physical quantity varies with load and with details the hypothesis does not fix. The complete statement is a *predictive distribution* for the outcome y_m of candidate m under each hypothesis, $p(y_m \mid H_i, y)$, conditioned on the record y because the record has already narrowed the parameters of each hypothesis. Averaged over the current posterior, it gives the predictive distribution of the outcome before the measurement is taken,

$$p(y_m \mid y) = \sum_i p(y_m \mid H_i, y)\ \Pr(H_i \mid y). \tag{19.2}$$

For the rack airflow, an illustrative pair of predictive distributions is

$$\dot{m}_{\text{air}} \mid H_1 \sim \mathcal{N}(1.5,\ 0.2^2)\,\text{kg/s}, \qquad \dot{m}_{\text{air}} \mid H_2 \sim \mathcal{N}(3.0,\ 0.3^2)\,\text{kg/s}. \tag{19.3}$$

The low-airflow value is consistent with the heat balance $\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, \Delta T$ of Chapter 5: an illustrative 30 kW rack at 3.0 kg/s heats its air by $30/(3.0 \times 1.0) = 10$ K, and at 1.5 kg/s by 20 K. The two means differ by 1.5 kg/s; the standard deviation of their difference is $\sqrt{0.2^2 + 0.3^2} = 0.36$ kg/s, so the means are about four standard deviations apart. A useful measurement is one whose predicted outcomes differ under the competing hypotheses, and this one does.

For the hot-aisle temperature the predictions are

$$T_{\text{hot}} \mid H_1 \sim \mathcal{N}(38,\ 2^2)^\circ\text{C}, \qquad T_{\text{hot}} \mid H_2 \sim \mathcal{N}(39,\ 2^2)^\circ\text{C}, \tag{19.4}$$

with $\mathcal{N}(38.5,\ 2^2)^\circ$C added here for H_3 as an illustrative middle value. The means differ by 1 K and the standard deviation of the difference is $\sqrt{2^2 + 2^2} = 2.8$ K: the means are a third of a standard deviation apart. Whatever this sensor reads, both hypotheses could have produced it.

Figure 19.2 draws both cases. The left panel has the rack airflow in kilograms per second on the horizontal axis and the probability density on the vertical axis; the right panel has the hot-aisle temperature in degrees Celsius. In each panel the solid red curve is the prediction under H_1 and the dashed blue curve under H_2; the dotted gray curve is H_3, which in the airflow panel lies exactly on the H_2 curve. The shaded area is the overlap, the probability mass that both hypotheses share. In the airflow panel the two curves cross near 2.1 kg/s and the overlap is 0.3 % of the area under either curve, too small to see at this scale; in the temperature panel it is 80 %. The left panel is a good measurement for separating H_1 from H_2; the right panel is a poor one, however precisely each reading is taken.

Principle 19.1. *A measurement separates hypotheses only to the extent that their predictive distributions for it differ. Two hypotheses that predict nearly the same value for a quantity cannot be told apart by measuring that quantity, however accurately.*

19.2 Information gain

This section turns the visual comparison of predictive distributions into a number. It defines the entropy of a diagnosis, the expected information gain of a measurement, works the ambiguous rack through with the rack visit and with the hot-aisle sensor, and then sets the gain against the cost of obtaining it.

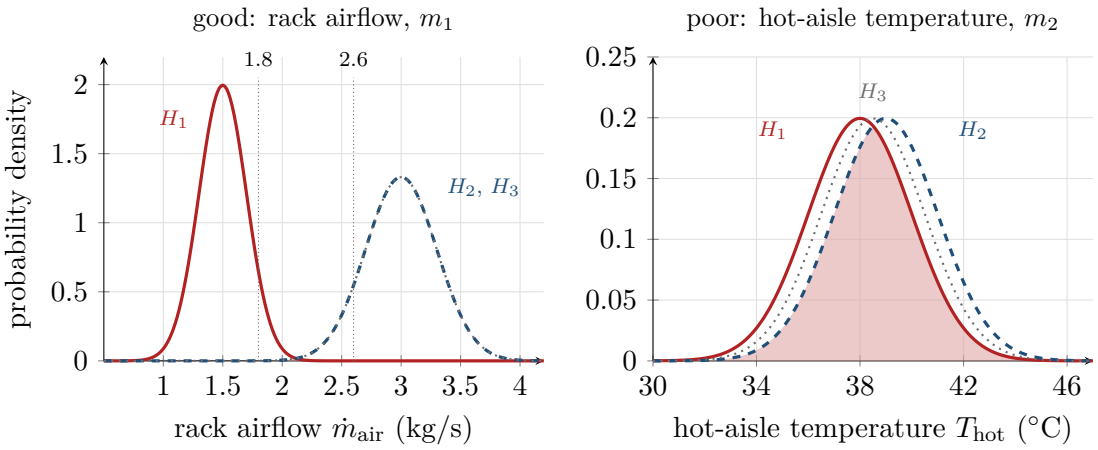


Figure 19.2: Predictive distributions of two candidate measurements, illustrative. Left: rack airflow, $\mathcal{N}(1.5, 0.2^2)$ under H_1 (solid) and $\mathcal{N}(3.0, 0.3^2)$ kg/s under H_2 (dashed) and H_3 (dotted, identical to H_2); the overlap, near 2.1 kg/s, is 0.3 %; dotted verticals mark the 1.8 and 2.6 kg/s thresholds of Section 19.5.5. Right: hot-aisle temperature, $\mathcal{N}(38, 2^2)$, $\mathcal{N}(39, 2^2)$ and $\mathcal{N}(38.5, 2^2)$°C; the overlap of H_1 and H_2 (shaded) is 80 %.

19.2.1 Entropy of a diagnosis

The uncertainty of a posterior over hypotheses is measured by its entropy,

$$\mathcal{H}(y) = -\sum_i \Pr(H_i \mid y)\, \log_2 \Pr(H_i \mid y), \tag{19.5}$$

in bits. Each term is the probability of a hypothesis times the "surprise" $-\log_2 \Pr(H_i \mid y)$ of learning that it is the true one, so the entropy is the average surprise. It is zero when one hypothesis has probability one and the others zero, and it is largest, $\log_2 n$ for n hypotheses, when all are equally likely. For three hypotheses the maximum is $\log_2 3 = 1.585$ bits. The entropy is high when the belief is spread over several hypotheses and low when one dominates.

For the posterior of equation (19.1), with $\log_2 0.45 = -1.152$, $\log_2 0.40 = -1.322$ and $\log_2 0.15 = -2.737$,

$$\mathcal{H}(y) = 0.45 \times 1.152 + 0.40 \times 1.322 + 0.15 \times 2.737 = 0.518 + 0.529 + 0.411 = 1.458 \text{ bits}. \tag{19.6}$$

That is 92 % of the three-hypothesis maximum: the diagnosis is close to as uncertain as three hypotheses allow. A posterior of $(0.90, 0.07, 0.03)$, by comparison, has an entropy of 0.557 bits.

19.2.2 Expected information gain

A measurement m with outcome y_m changes the posterior to $\Pr(H_i \mid y, y_m)$ by Bayes' rule,

$$\Pr(H_i \mid y, y_m) = \frac{p(y_m \mid H_i, y)\, \Pr(H_i \mid y)}{p(y_m \mid y)}, \tag{19.7}$$

and so changes the entropy to $\mathcal{H}(y, y_m)$. The outcome is not known in advance, so the value of the measurement is assessed on average over the outcomes it might return, weighted by the predictive distribution of equation (19.2). The *expected information gain* is

$$\mathrm{IG}(m) = \mathcal{H}(y) - \mathbb{E}_{y_m}\left[\mathcal{H}(y, y_m)\right] = \mathcal{H}(y) - \int p(y_m \mid y)\, \mathcal{H}(y, y_m)\, \mathrm{d}y_m, \tag{19.8}$$

the diagnostic uncertainty the measurement is expected to remove. For a measurement with a finite set of outcome classes the integral is a sum.

Three properties make IG a sound criterion. First, it is never negative: a particular outcome can increase the entropy (a unexpected result can make a confident diagnosis less confident), but on average a measurement cannot increase uncertainty. Second, the posteriors after the possible outcomes, weighted by their predictive probabilities, average back to the current posterior,

$$\sum_{\text{outcomes}} p(y_m \mid y)\, \Pr(H_i \mid y, y_m) = \Pr(H_i \mid y), \tag{19.9}$$

so a list of hoped-for posteriors that does not satisfy this identity is not a valid prediction; this check is used in the worked example below. Third, $\mathrm{IG}(m) = 0$ exactly when every hypothesis predicts the same distribution for y_m: then equation (19.7) returns the prior for every outcome. The principle of Section 19.1.3 is this third property stated in words.

19.2.3 A worked example: the rack visit

Section 19.1.2 found that the rack airflow m_1 alone separates H_1 from H_2 and H_3 but not H_2 from H_3. A technician who goes to the rack with a flow hood can, at the same visit, hold a calibrated probe at the inlet (m_5). The combined visit has three outcome classes: A, low airflow; B, normal airflow with a hot inlet confirmed by the probe; C, normal airflow with a normal inlet temperature at the probe. Table 19.2 gives the probability of each class under each hypothesis. The entries are illustrative and deliberately short of zero and one: a handheld flow reading has an error of several percent, a recirculation pattern can weaken at the moment of the visit, and a biased sensor can coincide with a mildly warm rack.

Table 19.2: Likelihood of each outcome class of the rack visit (airflow and calibrated inlet probe) under each hypothesis, and the predictive probability of each class under the posterior $(0.45, 0.40, 0.15)$. A: low airflow; B: normal airflow, inlet hot; C: normal airflow, inlet normal. Illustrative.

	outcome A	outcome B	outcome C
H_1 low airflow, Pr = 0.45	0.92	0.07	0.01
H_2 recirculation, Pr = 0.40	0.08	0.85	0.07
H_3 sensor bias, Pr = 0.15	0.09	0.11	0.80
predictive probability $p(y_m \mid y)$	0.46	0.39	0.15

The predictive probability of each class is the sum of its column weighted by the posterior, equation (19.2): for A, $0.45 \times 0.92 + 0.40 \times 0.08 + 0.15 \times 0.09 = 0.4140 + 0.0320 + 0.0135 = 0.4595 \approx 0.46$; for B, $0.0315 + 0.3400 + 0.0165 = 0.3880 \approx 0.39$; for C, $0.0045 + 0.0280 + 0.1200 = 0.1525 \approx 0.15$. Bayes' rule, equation (19.7), divides each product by its column total and gives the three possible posteriors, rounded to two decimals:

$$\begin{aligned} &\text{A:} \quad (0.4140,\ 0.0320,\ 0.0135)/0.4595 = (0.90,\ 0.07,\ 0.03), \\ &\text{B:} \quad (0.0315,\ 0.3400,\ 0.0165)/0.3880 = (0.08,\ 0.88,\ 0.04), \\ &\text{C:} \quad (0.0045,\ 0.0280,\ 0.1200)/0.1525 = (0.03,\ 0.18,\ 0.79). \end{aligned} \tag{19.10}$$

A low airflow makes H_1 dominant; a normal airflow with a confirmed hot inlet makes H_2 dominant; a normal airflow with a normal probe reading makes the sensor bias H_3 the leading explanation. The identity of equation (19.9) holds: for H_1, $0.46 \times 0.90 + 0.39 \times 0.08 + 0.15 \times 0.03 = 0.45$, and the same check returns 0.40 and 0.15 for the other two.

The entropies of the three posteriors are

$$\begin{aligned} \mathcal{H}_{\mathrm{A}} &= 0.137 + 0.269 + 0.152 = 0.557 \text{ bits}, \\ \mathcal{H}_{\mathrm{B}} &= 0.292 + 0.162 + 0.186 = 0.640 \text{ bits}, \\ \mathcal{H}_{\mathrm{C}} &= 0.152 + 0.445 + 0.269 = 0.866 \text{ bits}, \end{aligned}$$

where each term is $-\Pr \log_2 \Pr$ for one hypothesis. The expected entropy after the visit is their average weighted by the predictive probabilities,

$$\mathbb{E}_{y_m}\left[\mathcal{H}(y, y_m)\right] = 0.46 \times 0.557 + 0.39 \times 0.640 + 0.15 \times 0.866 = 0.636 \text{ bits}, \tag{19.11}$$

and the expected information gain is $\mathrm{IG} = 1.458 - 0.636 = 0.82$ bits, a 56 % expected reduction of the diagnostic uncertainty.

The hot-aisle temperature m_2 is the contrast. A typical reading of 37.5°C has densities proportional to $e^{-0.25^2/2} = 0.969$ under H_1, $e^{-0.75^2/2} = 0.755$ under H_2 and $e^{-0.5^2/2} = 0.882$ under H_3 (each exponent is the squared distance from the mean in units of the 2 K standard deviation, halved). Multiplied by the posterior and normalized, they give $(0.50, 0.35, 0.15)$: the belief barely moves, and its entropy falls only from 1.458 to 1.441 bits. Integrating over all readings with equation (19.8) gives an expected information gain of 0.04 bits, 2.6 % of the current uncertainty; the value is computed numerically by the chapter's data script, `scripts/ch19_data.py`.

Figure 19.3 draws the comparison. Each group of three bars is one posterior: the probability of H_1 (red), H_2 (blue) and H_3 (gray), on a vertical axis from zero to one. The first group is the current posterior; the next three are the posteriors after outcomes A, B and C of the visit, with their predictive probabilities in the axis labels; the last is the posterior after a 37.5°C hot-aisle reading. The entropy of each posterior is printed above its group. Every outcome of the visit leaves one hypothesis clearly ahead; the hot-aisle reading leaves the picture as it was.

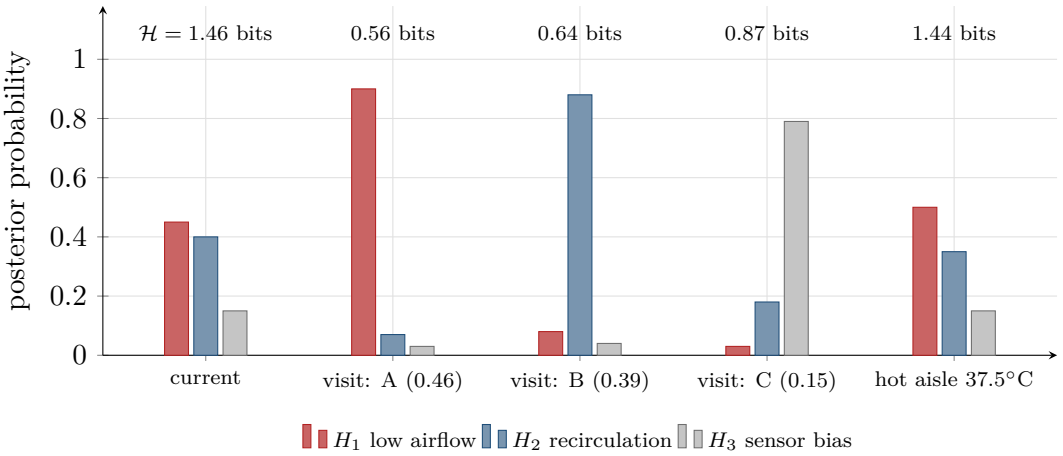


Figure 19.3: Posterior over the three hypotheses before and after a measurement, illustrative. Groups 2 to 4: the three outcome classes of the rack visit (Table 19.2), with predictive probabilities in parentheses; expected entropy 0.64 bits, expected information gain 0.82 bits. Group 5: a typical hot-aisle reading; expected information gain of that sensor 0.04 bits.

The visit is not perfect: each of its outcomes leaves 0.56 to 0.87 bits of uncertainty, because Table 19.2 allows for instrument error and intermittent faults. The airflow reading alone, with the sharper predictions of equation (19.3), would resolve H_1 almost completely but leave H_2 and H_3 at their current ratio of 0.40 to 0.15, that is $(0.73, 0.27)$ among themselves, whenever the airflow is normal. The reference probe is what splits that remainder.

19.2.4 Information and cost

Information is not the only criterion. The rack visit may need technician access to a live aisle, a flow hood that has to be fetched, and, if the measurement is taken inside the rack, a short shutdown of the servers. A slightly less informative measurement that is already in the telemetry may be preferable. In Table 19.1 the server fan speed m_4 costs nothing: it separates H_3 from

the other two, as the probe does, and it can be read before anyone is dispatched. If it shows normal fan speeds, the sensor-bias hypothesis rises sharply and the visit may be unnecessary; if it shows high fan speeds, H_3 falls and the visit is then targeted at the split between H_1 and H_2.

The comparison therefore needs a net value,

$$\text{net value}(m) = \text{information value}(m) - \kappa(m), \tag{19.12}$$

where $\kappa(m)$ is the cost of measurement m. The cost has several parts: the purchase of a sensor, its installation, any downtime of IT equipment while it is installed or used, the risk of performing the test (a technician in a hot aisle, a cable disturbed inside a rack), and the delay before the result is available, during which the facility operates on the current, uncertain belief. Information gain is in bits and cost is in money or time, so equation (19.12) needs an exchange rate between the two. The next section supplies it: it values information by what it changes in the decision, in the same units as the cost.

19.3 Decision value

Information gain measures how much a measurement teaches; it does not measure whether what it teaches matters. This section values information by the action it changes, works an example with a loss table, and states the general decision-theoretic formula.

19.3.1 When information does not change the action

Suppose a rack is overheating and two hypotheses remain, a degraded server fan H_{fan} and a blocked filter H_{filter}, each with probability 0.5. The entropy is one bit, the maximum for two hypotheses, and a test that resolves the question perfectly has an information gain of one bit. Both hypotheses, however, require the same immediate action: move load away from the rack now. Separating them does not change that action. Whichever hypothesis the test confirms, the operator does the same thing, the loss incurred is the same, and the test has no immediate operational value. It may still have value later, for the maintenance work order, but that is a different decision.

Information is therefore valued by whether it changes the action. A measurement can be scientifically interesting, reduce entropy by a full bit, and still have no value for the decision at hand.

19.3.2 Expected value of information

The expected value of information is

$$\text{VOI} = \text{expected loss without the measurement} - \text{expected loss after it}, \tag{19.13}$$

where "after it" means: averaged over the outcomes the measurement might return, with the best action chosen separately for each outcome. If the best action does not depend on the outcome, both terms are equal and $\text{VOI} = 0$, as in the fan-and-filter example.

Take an alarm with two remaining hypotheses: a sensor fault H_{s} with $\Pr(H_{\text{s}} \mid y) = 0.55$, and a cooling failure H_{c} with $\Pr(H_{\text{c}} \mid y) = 0.45$. Two actions are available: shed load from the affected racks immediately, a_{shed}, or continue operating, a_{cont}. Table 19.3 gives illustrative losses in thousands of dollars. Shedding costs 20 under either hypothesis (interrupted workloads and service penalties). Continuing costs nothing if the alarm is a sensor fault and 200 if it is a cooling failure, because the racks then overheat and trip.

Without further information the expected losses are 20 for shedding and $0.55 \times 0 + 0.45 \times 200 = 90$ for continuing. Because the cooling failure is severe, the current best action is to reduce load immediately, with an expected loss of 20. Now suppose a second, independent temperature

Table 19.3: Loss of each action under each hypothesis, and expected loss under the current posterior $(0.55, 0.45)$, in k$. Illustrative.

Action	H_{s} sensor fault	H_{c} cooling failure	expected loss
a_{shed}: reduce load now	20	20	20
a_{cont}: continue	0	200	90
best action given the hypothesis	a_{cont}, 0	a_{shed}, 20	–

sensor near the rack can be read in about 30 seconds and resolves the ambiguity: if it agrees with the alarming sensor, the cooling failure is confirmed; if it reads normal, the alarm is a sensor fault. After the reading the operator continues under H_{s}, with loss 0, and sheds under H_{c}, with loss 20. The expected loss after the measurement is

$$0.55 \times 0 + 0.45 \times 20 = 9, \qquad \mathrm{VOI} = 20 - 9 = 11 \text{ k\$}. \tag{19.14}$$

The measurement avoids an unnecessary load shed in the 55 % of cases where the alarm is a sensor fault, and that is worth $0.55 \times 20 = 11$ thousand dollars on average. Its cost is 30 seconds of delay, which is negligible when the thermal time constants of the room are minutes (Chapter 8). The decision value is high.

Figure 19.4 shows how this value depends on the current belief. The horizontal axis is the probability of the cooling failure, $\Pr(H_{\mathrm{c}} \mid y)$, from zero to one. The solid green curve, read on the left axis in thousands of dollars, is the value of the resolving measurement; the dashed blue curve, read on the right axis in bits, is the entropy of the two-hypothesis posterior. Below a probability of 0.1 the best action without information is to continue, with expected loss $200 \Pr(H_{\mathrm{c}} \mid y)$; above it, to shed, with loss 20. The value of information is the gap between the better of those two and the expected loss with perfect information, $20 \Pr(H_{\mathrm{c}} \mid y)$. It peaks at 18 thousand dollars exactly at the probability 0.1 where the best action switches, and falls to zero at both ends, where the action is already clear. The entropy peaks at 0.5 instead. The two measures agree that a certain belief needs no measurement, but they disagree about where a measurement is most needed: information is most needed where the belief is most spread; decision value is greatest where the belief is closest to the point at which the decision changes.

Principle 19.2. *Information has decision value only if some outcome of the measurement would change the best action. The value is greatest when the current belief is near the point at which the action switches, not necessarily when the belief is most uncertain.*

19.3.3 The decision-theoretic formulation

The example generalizes. Let a range over the available actions and let $\ell(a, H_i)$ be the loss of action a when H_i is true. With the current information, here the measurement record y, the best achievable expected loss is

$$\ell^*(y) = \min_a \ \mathbb{E}[\ell(a, H) \mid y] = \min_a \sum_i \ell(a, H_i) \ \Pr(H_i \mid y). \tag{19.15}$$

If a candidate measurement m is taken and returns y_m, the posterior becomes $\Pr(H_i \mid y, y_m)$ and the best expected loss becomes

$$\ell^*(y, y_m) = \min_a \sum_i \ell(a, H_i) \ \Pr(H_i \mid y, y_m). \tag{19.16}$$

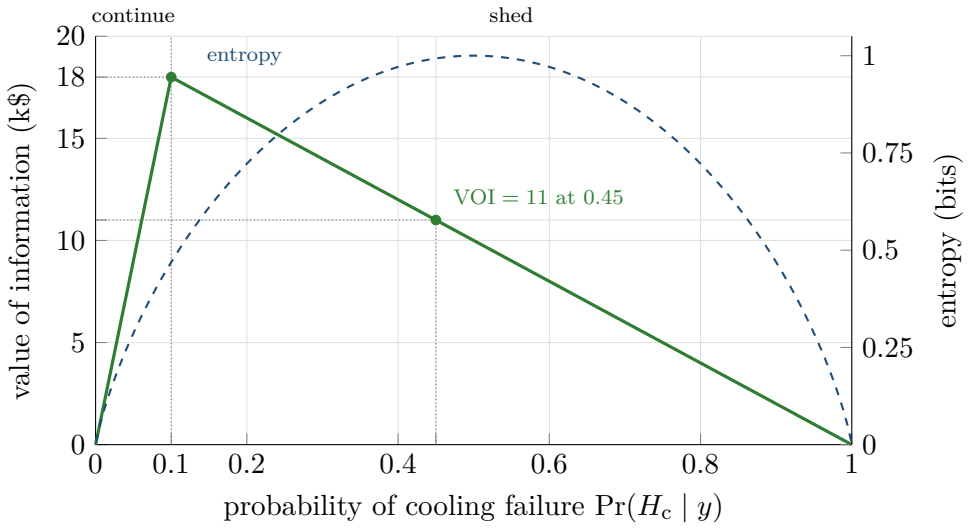


Figure 19.4: Value of a resolving measurement (solid, left axis) and entropy of the posterior (dashed, right axis) against the probability of a cooling failure, for the losses of Table 19.3: $\mathrm{VOI}(p) = \min(20,\ 200p) - 20p$ k\$ and $\mathcal{H}(p) = -p\log_2 p - (1-p)\log_2(1-p)$, with $p = \Pr(H_c \mid y)$. Illustrative. The value peaks at 18 k\$ at $p = 0.1$ (dotted), where the best action without information switches; the labels above the plot name that action on each side.

The value of the measurement is the expected reduction in decision loss,

$$\mathrm{VOI}(m) = \ell^*(y) - \mathbb{E}_{y_m}\left[\ell^*(y, y_m)\right], \tag{19.17}$$

with the expectation taken over the predictive distribution $p(y_m \mid y)$ of equation (19.2). The next-best measurement maximizes the net value $\mathrm{VOI}(m) - \kappa(m)$, and none is worth taking if every net value is negative.

Equation (19.17) has the same structure as the information gain of equation (19.8): a current quantity minus its expectation after the measurement. The entropy is replaced by the best achievable loss. Like the information gain, the value of information is never negative, because choosing the action after seeing the outcome can never do worse on average than choosing it before. When the measurement resolves the hypotheses completely, the expected loss after it is $\sum_i \Pr(H_i \mid y)\min_a \ell(a, H_i)$, the loss with perfect information, and the value is the largest any measurement can have; the example of equation (19.14) is that case. An imperfect measurement, such as the rack visit of Table 19.2, has a smaller value, computed with the same formula over its outcome classes.

19.4 Active tests and experiment design

A measurement observes the system as it is. A test changes the system slightly and observes its response. This section treats tests as measurements with a cost and a safety constraint, then the offline version of the same problem, the placement of sensors, and the design of experiments for parameters rather than hypotheses.

19.4.1 Measurement versus intervention

Some hypotheses predict the same steady readings and differ only in how the system responds to a change. For those, no passive measurement separates them, but a small intervention can: change a control input slightly, for example raise a CRAH fan command u_{fan} from 50 % to 60 %, and observe the thermal and airflow response. This is an *active diagnostic test*.

Consider two hypotheses for an overheating zone: H_1, degradation of the CRAH fan (a failing motor or a slipping belt), and H_2, recirculation. Raise the fan speed. Under H_2 the fan is healthy, and by the fan law of Chapter 6 its flow scales with speed: the delivered airflow rises by about $60/50 - 1 = 20\,\%$. The inlet temperature nevertheless stays elevated, because the hot air at the inlet arrives over the top of the rack and more supply into the cold aisle does little to stop it. Under H_1 the degraded fan cannot respond: the delivered airflow changes little, here an illustrative $3\,\%$, and the inlet temperature also stays elevated. The separating signal is therefore the airflow response, and the temperature response confirms that the problem is not a simple shortage of supply air, which would show a temperature drop.

Figure 19.5 draws the test. Both panels have time in seconds after the step on the horizontal axis. The left panel shows the delivered airflow as a percentage of its value before the step; the right panel shows the inlet temperature of the affected rack in degrees Celsius. The dashed blue curves are the response predicted under H_2, the solid red curves under H_1. The dotted gray curve in the right panel is a reference that neither hypothesis predicts: the temperature drop a healthy fan would produce in a zone whose only problem is too little supply air. The airflow panel separates the hypotheses within a minute; the temperature panel shows that both hypotheses leave the inlet hot.

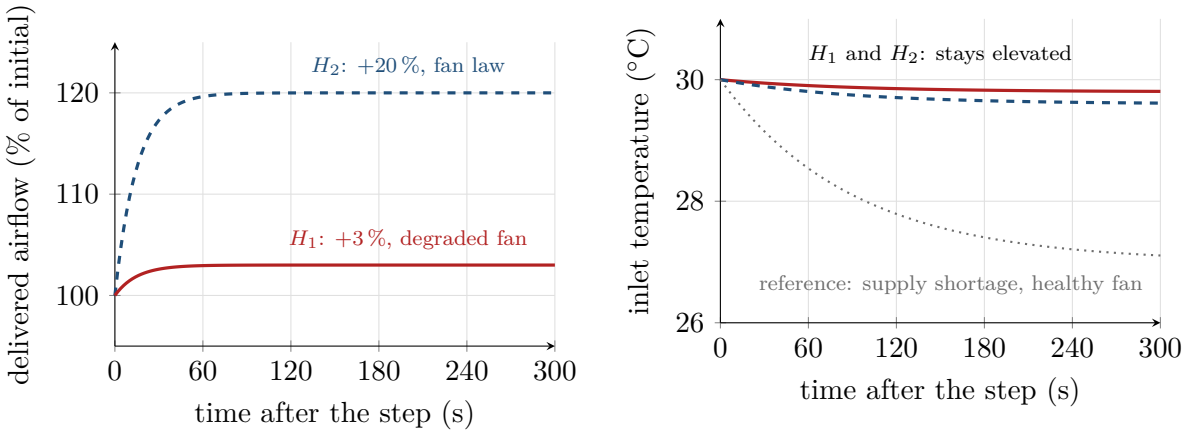


Figure 19.5: An active diagnostic test: the CRAH fan command steps from 50 % to 60 % at $t = 0$. Left: delivered airflow, $100 + \Delta(1 - e^{-t/15\,\mathrm{s}})$ with $\Delta = 20\,\%$ under H_2 (healthy fan) and 3 % under H_1 (degraded fan). Right: inlet temperature, $30 - \delta(1 - e^{-t/90\,\mathrm{s}})$°C with $\delta = 0.2\,\mathrm{K}$ (H_1), 0.4 K (H_2) and, for reference, 3 K for a zone short of supply air. Illustrative.

19.4.2 Controls as experiments, under safety constraints

The fan-speed test treats a control input as an experiment. In normal operation a control input u is chosen for performance: to hold a temperature at its setpoint at the lowest energy (Chapter 10). A test input u_{test} is chosen for information: to make the hypotheses predict different responses. Choosing control actions that serve both purposes is *dual control*: control for performance plus control for information. A small, well-placed perturbation inside the normal operating range can teach a great deal while costing almost nothing in performance.

A test is not free of risk. The perturbation must keep every rack inlet temperature T_i below its limit, the facility power below its limit, and the IT service within its commitments, and it must do so under every hypothesis that is still plausible, not only under the most probable one. Choosing the next test is therefore a constrained optimization problem,

$$\max_{u_{\text{test}}} \ \mathrm{VOI}(u_{\text{test}}) - \kappa(u_{\text{test}}) \quad \text{subject to} \quad T_i(t;\ u_{\text{test}}, H_j) \le T_{\max}, \quad P_{\text{facility}}(t;\ u_{\text{test}}, H_j) \le P^{\max}, \tag{19.18}$$

for every rack i, every time t of the test and every hypothesis H_j with non-negligible posterior probability, together with the service constraints of the IT load. The constraint is written for each hypothesis because the test is run before the hypothesis is known: a step that is safe under H_2 may be unsafe under H_1 if, for example, the step is a reduction of fan speed and the fan is already weak. Raising a fan speed is a natural test for this reason: it moves every plausible hypothesis toward safety, at the cost of a little fan power.

19.4.3 Sensor placement

Sensor placement is the offline version of the same problem. A new data hall has a budget for 20 temperature sensors. The naive layout spreads them evenly, one per few racks along the centers of the cold aisles. The better layout places them to maximize information about the operational questions that matter, each sensor where the competing explanations of a likely problem predict different readings.

If the main uncertainty in the hall is recirculation, sensors only at cold-aisle centers are poor. The air there is close to the supply temperature under almost every hypothesis, so those sensors predict nearly the same reading whether or not recirculation is present, and they are nearly redundant with each other. Recirculated air shows first at the tops of racks, where exhaust spills over, and at the ends of rows, where it wraps around. Better locations are the inlets at the top of the racks, the boundaries between hot and cold aisles at row ends, and any suspected recirculation path such as a gap in containment or a missing blanking panel. The general rule is to tie instrumentation to model sensitivity and to the inference goals: a sensor belongs where the derivative of its reading with respect to the uncertain quantity is large, and where that derivative differs from those of the sensors already installed.

19.4.4 Exposing topology

Tests can also reveal topology, the connection pattern of the system. Suppose it is not known whether CRAH 2 mainly serves row A or row B: the drawings are out of date and the underfloor flow path is uncertain. Static correlations in historical telemetry are a poor guide, because both rows and the CRAH respond together to the same daily load cycle and the same supply temperature; everything correlates with everything. A deliberate step breaks that common driver. Raise the CRAH 2 fan speed N_{fan} by 10 % (by the fan law, about 10 % more air) and record the row temperatures $T_{\text{row A}}(t)$ and $T_{\text{row B}}(t)$. The row with the stronger and earlier response is the row CRAH 2 serves.

Figure 19.6 shows an illustrative result. The horizontal axis is minutes from the step; the vertical axis is the change of each row's mean inlet temperature from its value before the step. Row A (solid) falls by 0.8 K with a time constant of about 3 minutes after a half-minute transport delay; row B (dashed) falls by only 0.1 K. Row A is served by CRAH 2. Both changes are small, well inside the normal range of either row, which is what makes the test safe; the information lies in the contrast between the two responses, not in their size.

19.4.5 Experiments for parameter estimation

Experiments are also designed to estimate parameters rather than to choose between hypotheses. Suppose the thermal resistance R_{th} and capacitance C_{th} of a zone, the two parameters of the first-order model $\tau = R_{\text{th}}C_{\text{th}}$ of Chapter 8, are poorly identified from routine data. Routine data are close to steady: the steady temperature rise is $R_{\text{th}}\dot{Q}$, which fixes R_{th} and says nothing about C_{th}. A short load pulse, a fan-speed step or a chilled-water setpoint change excites the dynamics and makes C_{th} visible.

An illustrative case shows why. Two parameter pairs share $R_{\text{th}} = 0.5$ K per kilowatt but have $C_{\text{th}} = 600$ and 1200 kJ/K, so $\tau = 0.5 \times 600 = 300$ s and 600 s. Under a constant 20 kW both

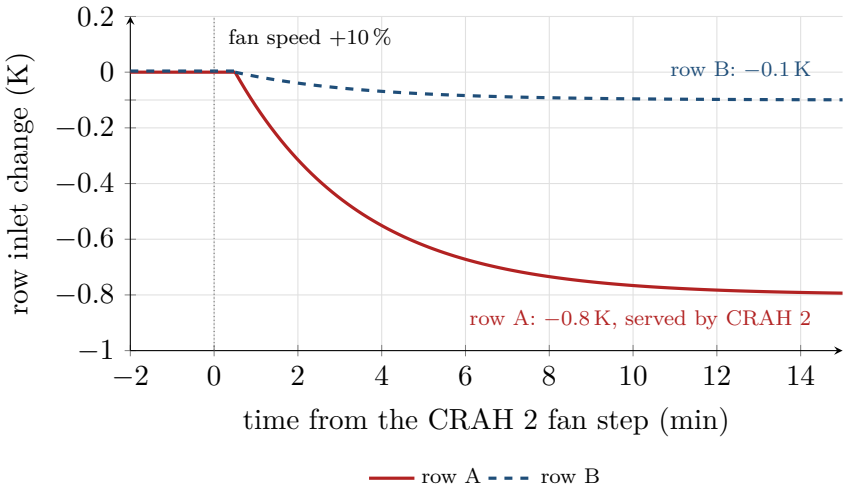


Figure 19.6: A topology test, illustrative: response of the two rows' mean inlet temperatures to a 10 % step of the CRAH 2 fan speed at $t = 0$. Row A: $-0.8(1 - e^{-(t-0.5)/3})$ K; row B: $-0.1(1 - e^{-(t-0.5)/3})$ K, with t in minutes and a half-minute transport delay. The stronger response identifies the row the unit serves.

settle at the same rise, $0.5 \times 20 = 10$ K, and no steady record separates them. Under a 20 kW step they separate at once: the initial slopes are $\dot{Q}/C_{\text{th}} = 20/600 = 0.033$ K/s, 2 K per minute, and 1 K per minute; after five minutes the rises are $10(1-e^{-1}) = 6.3$ K and $10(1-e^{-0.5}) = 3.9$ K. The best experiment is the one that gives competing parameter combinations different dynamic signatures, as large as the safety constraints of Section 19.4.2 allow. Its systematic form is *optimal experiment design*; Chapter 15 treats the identifiability question it answers.

19.4.6 Fisher information and diverse measurements

For a parameter vector θ and a measurement $y = h(\theta) + v$ with Gaussian noise v of covariance Σ, the information the measurement carries about θ is the Fisher information matrix

$$\mathcal{I}(\theta) = J^{\mathsf{T}}\,\Sigma^{-1} J, \qquad J = \frac{\partial h}{\partial \theta}, \tag{19.19}$$

where J is the sensitivity matrix, one row per measurement and one column per parameter. Independent measurements add their information matrices, and a prior with covariance Σ_0 adds Σ_0^{-1}. The inverse of the total information approximates the posterior covariance of θ, whose one-standard-deviation contour is an ellipse (an ellipsoid in more dimensions) with area proportional to $1/\sqrt{\det \mathcal{I}}$. The *D-optimal* criterion chooses the measurement or experiment that maximizes $\det \mathcal{I}(\theta)$: it favors measurements that are sensitive to the uncertain parameters and penalizes redundancy, because a new row of J parallel to existing rows adds information only along a direction that is already constrained.

An illustrative example makes this concrete. Take $\theta = (\dot{m}_{\text{air}}, \alpha)$, the airflow and the recirculation fraction of one rack, around $\dot{m}_{\text{air}} = 3.0$ kg/s and $\alpha = 0.20$, for a 30 kW rack whose air heats by $\Delta T = 10$ K. Combining the mixing rule of Chapter 6 with the rack heat balance gives the inlet temperature $T_{\text{in}} = T_{\text{supply}} + \alpha\,\Delta T/(1-\alpha)$, with $\Delta T = \dot{Q}_{\text{rack}}/(\dot{m}_{\text{air}} c_{p,\text{air}})$, so its sensitivities are

$$\frac{\partial T_{\text{in}}}{\partial \dot{m}_{\text{air}}} = -\frac{\alpha}{1-\alpha}\,\frac{\Delta T}{\dot{m}_{\text{air}}} = -0.83\ \text{K per kg/s}, \qquad \frac{\partial T_{\text{in}}}{\partial \alpha} = \frac{\Delta T}{(1-\alpha)^2} = 15.6\ \text{K}. \tag{19.20}$$

The rack's inlet sensor (noise 0.5 K) and a prior with standard deviations 0.6 kg/s and 0.10 give the current belief: $\dot{m}_{\text{air}}$ known to ± 0.57 kg/s, barely better than the prior, and α to

± 0.041, with the two strongly correlated (correlation 0.67), because the sensor constrains only the combination $-0.83\,\dot{m}_{\text{air}} + 15.6\,\alpha$. Adding an identical inlet sensor beside the first, with independent noise, doubles the information along that same combination: $\det \mathcal{I}$ rises from about 3.3×10^3 to 6.2×10^3, α improves to ± 0.037 and $\dot{m}_{\text{air}}$ hardly improves, from ± 0.574 to $\pm 0.573\,\text{kg/s}$. Adding instead a direct airflow measurement with noise 0.1 kg/s, sensitivity row $(1, 0)$, raises $\det \mathcal{I}$ to 1.1×10^5, a factor of 34: $\dot{m}_{\text{air}}$ becomes known to $\pm 0.10\,\text{kg/s}$ and, through the correlation, α to ± 0.031. Without the prior, the data alone make the point sharply: any number of identical temperature rows gives a matrix of rank one, with determinant zero, while one temperature row and one airflow row give a full-rank matrix.

Figure 19.7 draws the three beliefs. The horizontal axis is the rack airflow $\dot{m}_{\text{air}}$ and the vertical axis the recirculation fraction α; each closed curve is a one-standard-deviation ellipse. The current belief (solid gray) is a long tilted ellipse, narrow across the direction the temperature sensor constrains and long along the direction it cannot see. The identical second sensor (dashed blue) narrows it a little across that direction and leaves its length unchanged. The airflow measurement (solid green) cuts the length. Diverse measurements can be worth more than more of the same type.

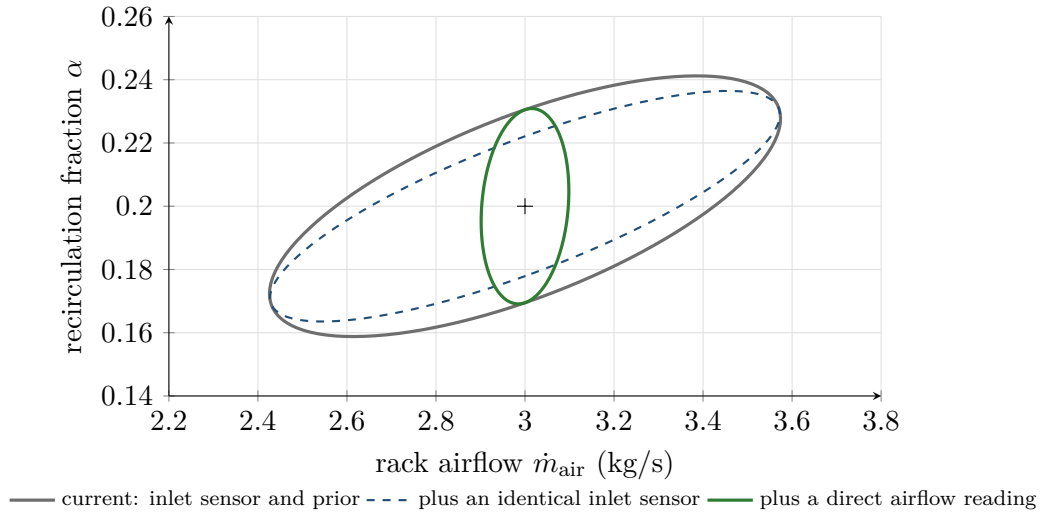


Figure 19.7: One-standard-deviation ellipses of the belief about rack airflow and recirculation fraction, from the inverse Fisher information $(\Sigma_0^{-1} + \sum J^{\mathsf{T}} \Sigma^{-1} J)^{-1}$, centered on $(3.0\,\text{kg/s}, 0.20)$ (cross). Prior standard deviations 0.6 kg/s and 0.10; inlet-temperature row $(-0.83, 15.6)$ with noise 0.5 K; airflow row $(1, 0)$ with noise 0.1 kg/s. Computed by `scripts/ch19_data.py`. Illustrative.

19.5 Sequential and question-specific measurement

The next-best measurement is not a fixed property of a facility. It depends on what is already known, on when to stop, and on which question the measurement is meant to answer. This section treats measurement as a sequence, distinguishes the kinds of uncertainty a measurement can target, and shows how a recommendation, a temporary measurement and a capital purchase are each chosen and stated.

19.5.1 Value depends on current uncertainty

The value of a measurement depends on the uncertainty that remains when it is taken. In the example of Figure 19.7, the record so far has reduced the uncertainty of the recirculation

fraction from ±0.10 to ±0.041, by 59 %, and that of the airflow from ±0.60 to ±0.57 kg/s, by 4 %. Today the uncertainty is mostly in $\dot{m}_{\text{air}}$, and an airflow measurement has the highest value. After it is taken, airflow is known to about 3 % of its value (0.10 of 3.0 kg/s) while the recirculation fraction is still uncertain by about 15 % (0.031 of 0.20). The uncertainty has shifted to α, and a measurement that targets it, such as a local temperature map at the top of the rack and the aisle boundary, now has the higher value. A second airflow reading would add little.

The ranking of candidate measurements is therefore adaptive. It is recomputed after every result, and a list of "the most valuable sensors" fixed in advance is right only for the belief it was computed from.

19.5.2 The sequential loop and stop rules

Measurement is a sequential process. Starting from the record y_0 and the posterior $\Pr(H \mid y_0)$, the next-best measurement m_1 is chosen and taken, returning y_1; the posterior is updated to $\Pr(H \mid y_0, y_1)$; the next measurement m_2 is chosen from that posterior; and so on. The loop ends when one of three conditions holds: the uncertainty is low enough, the decision is clear, or no remaining measurement is worth its cost. This is *sequential experimental design*.

Figure 19.8 draws the loop. The boxes along the top are its four steps: infer the posterior from all the data so far, test the stop rule, choose the measurement or test with the largest net value VOI − κ subject to the safety constraints, and take it. The arrow returning along the bottom carries the new result back into the inference. The branch to the right leaves the loop when the stop rule is met, and the decision is then made on the current posterior.

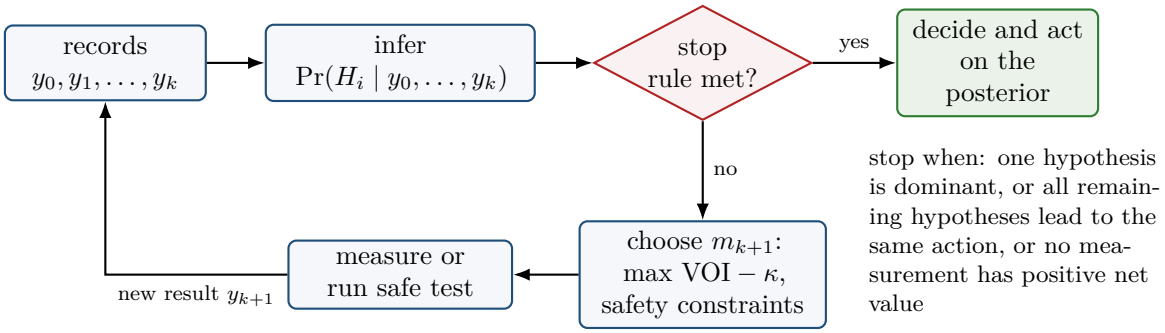


Figure 19.8: Sequential measurement. The posterior is inferred from all results so far; if the stop rule is not met, the measurement or safe test with the largest net value is chosen, taken, and its result is added to the record.

Stop rules. Knowing when enough information has been collected is part of the method. If $\Pr(H_1 \mid y) = 0.85$ and all the remaining hypotheses lead to the same maintenance action as H_1, more measurement is wasteful: by equation (19.17) every candidate has a value of zero for that decision, whatever its information gain, because no outcome can change the action. The three stop conditions of Figure 19.8 are the three ways the net value of every remaining candidate can become zero or negative: no uncertainty left to remove, no action left to change, or no measurement cheap enough.

19.5.3 Diagnosis, risk and decision uncertainty

A measurement can target three different uncertainties, and they are not always the same. *Diagnosis uncertainty* is the spread of the posterior over hypotheses, measured by the entropy $\mathcal{H}(y)$. *Risk uncertainty* is the spread of the predicted probability of an unsafe outcome, such as

$\Pr(T > T_{\max})$ for a rack inlet over the next hour, the risk measure of Chapter 17. *Decision uncertainty* is whether the best action is already clear, measured by the value of information.

A blocked filter and a degraded fan show the difference. Both reduce the same airflow by about the same amount, so they imply about the same future airflow and nearly the same thermal risk; illustratively, $\Pr(T > T_{\max})$ in the next hour is 0.08 under one and 0.09 under the other. Which of the two is present matters for maintenance, which component to replace, but hardly at all for immediate risk. If, within either hypothesis, the actual airflow is uncertain by $\pm 0.4\,\mathrm{kg/s}$, and that range spans risks from 0.01 to 0.25, then the airflow uncertainty dominates the risk. Measuring the airflow is then valuable for safe operation even though it does not diagnose the fault, and the filter pressure drop is valuable for the repair even though it hardly changes the risk.

The next-best measurement therefore depends on the goal: safe operation, repair planning or root-cause analysis. A recommendation that does not state which of these it serves cannot be checked.

19.5.4 Question-specific measurements and temporary sensors

Since value is defined with respect to a decision, the same facility has different next-best measurements for different questions. Table 19.4 lists three. Whether 500 kW of new IT load can be added safely depends mostly on the airflow the room can deliver to the affected racks, so the airflow measurement has the highest value. Which component needs maintenance depends on the pressure drop across the filter or the current drawn by the fan motor. Whether temperature sensor 4 is biased is answered by a temporary calibrated reference sensor beside it.

Table 19.4: Different questions about the same facility and their next-best measurements.

Question	Goal	Dominant uncertainty	Next-best measurement
A: can 500 kW be added safely?	safe operation, capacity	delivered airflow to the racks	rack airflow $\dot{m}_{\mathrm{air}}$
B: which component needs maintenance?	repair planning	filter or fan	filter pressure drop Δp, or fan motor current
C: is temperature sensor 4 biased?	root cause, data quality	sensor bias b	temporary calibrated reference sensor

Temporary sensors. Many of these measurements need not be permanent. A technician can measure airflow with a flow hood, temperature with a calibrated probe, differential pressure with a portable manometer and electrical current with a clamp meter, for an hour or a day. One temporary measurement that collapses a major ambiguity is cheaper than permanent instrumentation everywhere, and the sequential loop of Figure 19.8 tells where to send the technician.

19.5.5 Stating the recommendation and its expected outcomes

A recommendation states the measurement, its location and its expected value. For a rack with $\Pr(\text{low airflow}) = 0.48$, $\Pr(\text{sensor bias}) = 0.42$ and 0.10 left for recirculation, a diagnostic system might recommend: take one calibrated rack inlet temperature and airflow measurement at rack 12; expected posterior uncertainty reduction 75 %. The figure is an expected entropy reduction computed as in Section 19.2.3. With calibrated instruments whose reading falls in the matching outcome class with probability 0.94 and in each other class with 0.03, the entropy

of $(0.48, 0.10, 0.42)$ is 1.37 bits and the expected entropy after the measurement is 0.34 bits, a reduction of 1.02 bits or 75 %.

"Measure airflow" is not yet a useful recommendation. A useful one states the expected outcomes and what each would mean: if the airflow is below 1.8 kg/s, the low-airflow hypothesis becomes dominant; if it is above 2.6 kg/s, recirculation (or, with the probe, sensor bias) becomes much more likely. With the predictive distributions of equation (19.3), a reading of 1.8 kg/s is about 1.5×10^3 times more probable under H_1 than under H_2, and a reading of 2.6 kg/s about 10^6 times more probable under H_2 than under H_1; the dotted lines of Figure 19.2 mark the two thresholds. A reading between them is unlikely under either hypothesis (probability 0.07 under H_1 and 0.09 under H_2) and calls for a repeat reading before anything else.

Stating the outcomes in advance makes the recommendation auditable. The sentence "these hypotheses predict different values for X" can be checked by anyone with the model and the record, before the measurement is taken, and the result can then be compared with the prediction. A recommendation with stated expected outcomes is a falsifiable prediction, and that is the basis of trust in it.

19.5.6 Capital prioritization

The same calculation ranks permanent investments. Suppose a facility can fund one permanent instrumentation package: A, rack airflow sensors, \$100k; B, more temperature sensors, \$40k; or C, additional chilled-water flow metering, \$70k. For each, the value of information can be computed for each use: how much it improves capacity estimation, diagnosis and thermal-risk prediction, each expressed as an expected reduction of loss over the planning horizon. Table 19.5 gives illustrative values over five years.

Table 19.5: Value of information of three instrumentation packages over a five-year horizon, in k\$, by use. Illustrative values.

Package	cost	capacity	diagnosis	thermal risk	total VOI	VOI − cost
A: rack airflow sensors	100	120	90	80	290	190
B: more temperature sensors	40	20	30	25	75	35
C: chilled-water flow metering	70	90	15	30	135	65

In this illustration package A has the largest net value although it is the most expensive, because rack airflow is the quantity the facility infers least well (Section 19.4.6); package B is the cheapest but adds more of a kind of measurement the facility already has; package C serves mainly the capacity question at the plant level, where it closes the water-side heat balance. Two cautions apply. The values are not additive: once A is installed, the values of B and C change, as Section 19.5.1 showed. And the values depend on how often each decision recurs and what its losses are, so they are estimates to be revised as the facility's record grows.

19.6 Limits and reporting

The method of this chapter chooses among measurements on the assumption that the hypotheses and the model are right. This section treats the case in which they are not, gives the form of a practical report, and places the method in the progression from monitoring to experiment design.

19.6.1 The next-best model

Sometimes the bottleneck is not missing data but missing physics. If the model omits an airflow path, such as recirculation through a gap in the containment, then none of the hypotheses contains that path, and every hypothesis predicts nearly the same value for every candidate measurement near it. Every candidate then has a low value of information, while the residuals between the model and the data remain large and structured. That combination, low value of information everywhere and persistent residuals, is the signal that the next step is to improve the model, not to collect more data.

Measurements cannot fix model inadequacy. A model that fixes $\alpha = 0$ in a facility with significant recirculation will misread every inlet temperature that recirculation raises. One hundred more temperature sensors will show persistent residuals, and the inference will stay wrong: averaging more readings shrinks the random part of the error but not the systematic part the missing mechanism causes. The recirculation mechanism has to be added first.

Figure 19.9 quantifies this for an illustrative case. The horizontal axis is the number of temperature sensors n, on a logarithmic scale from 1 to 100; the vertical axis is the root-mean-square error of the inferred inlet temperature at an unmeasured rack, in kelvin. With an adequate model (solid green) the error falls as $1.2/\sqrt{n}$, from 1.2 K with one sensor to 0.12 K with one hundred. With the recirculation path missing (dashed red) the error is $\sqrt{1.5^2 + 1.2^2/n}$: it starts at 1.9 K and approaches the 1.5 K bias floor (dotted), which no number of sensors removes.

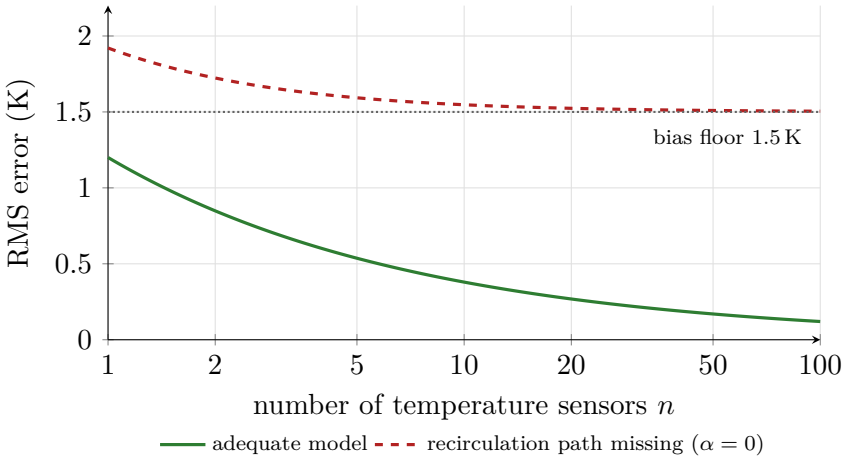


Figure 19.9: Error of an inferred inlet temperature against the number of temperature sensors, illustrative: $1.2/\sqrt{n}$ K for an adequate model and $\sqrt{1.5^2 + 1.2^2/n}$ K for a model without the recirculation path. More sensors shrink the random part of the error, not the bias of a missing mechanism.

19.6.2 A practical report

The output of the method, for an operator, is a short report. Table 19.6 gives an illustrative one for rack 18. It lists the unresolved hypotheses with their probabilities, the recommended measurement, the reason in the form of the predicted outcome under each hypothesis, the expected diagnostic value, the best alternative and its value, and the safety urgency, which says how long the measurement can wait.

Each line answers a question an operator would otherwise have to ask. The hypotheses and their probabilities say what is not known; the predicted ranges say why the recommended measurement resolves it and make the recommendation checkable; the alternative shows that

Table 19.6: A practical next-best-measurement report for rack 18, illustrative.

Item	Content
Unresolved hypotheses	reduced rack airflow 46 %; hot-air recirculation 41 %; sensor bias 13 %
Recommended measurement	direct rack airflow at rack 18
Why	predicted 1.4 to 1.8 kg/s under reduced airflow, 2.6 to 3.1 kg/s under recirculation
Expected diagnostic value	high
Alternative	hot-aisle temperature; moderate value
Safety urgency	thermal limit unlikely to be reached within 30 minutes

the choice was made against something; and the urgency line separates the diagnostic question from the safety question, which Section 19.5.3 showed are different. The entropy of the posterior $(0.46, 0.41, 0.13)$ is 1.43 bits, close to that of the example of Section 19.2.1.

19.6.3 From monitoring to experiment design

The method changes the kind of statement a monitoring system makes. Three levels can be distinguished:

1. "Sensor 14 is abnormal": a measured fact, an alarm.
2. "Airflow degradation is likely": an inferred statement, a diagnosis with a probability.
3. "The evidence cannot distinguish airflow degradation from recirculation; a rack airflow measurement would reduce the uncertainty most, and here is why": a statement about what is not known and how to learn it, an experiment design.

The third is scientific experiment design applied to operations. It keeps the book's distinction between measured, calculated and inferred quantities, and adds a fourth category: what is not yet known and which observation would make it known.

Faced with an uncertain inference, an operator can guess, act on the most probable hypothesis and accept being wrong a large part of the time; stop, and leave the ambiguity in place; or identify what information would resolve it. The method of this chapter is the third response. Its question is which observation, test or safe intervention would most improve the decision to be made, and its loop is: measure, infer, recognize the remaining uncertainty, choose the next measurement, and infer again.

19.7 What to retain

Definition 19.1 (Next-best measurement)**.** The next-best measurement is the observation, test or safe intervention whose expected reduction of decision loss, computed from the predictive distributions of its outcome under the competing hypotheses, most exceeds its cost, subject to safety constraints under every plausible hypothesis; it is chosen again after every result.

Table 19.7 lists the chapter's objects and the subsection that defines each.

Table 19.7: The objects of next-best measurement and where each is defined.

Object	Expression	Role	Defined in
Posterior	$\Pr(H_i \mid y)$	current belief over hypotheses	19.1.1
Predictive distribution	$p(y_m \mid H_i, y)$, eq. (19.2)	what a measurement would show	19.1.3
Entropy	$\mathcal{H}(y)$, eq. (19.5)	diagnostic uncertainty, bits	19.2.1
Information gain	$\mathrm{IG}(m)$, eq. (19.8)	expected entropy reduction	19.2.2
Measurement cost	$\kappa(m)$	purchase, installation, downtime, risk, delay	19.2.4
Value of information	$\mathrm{VOI}(m)$, eq. (19.17)	expected reduction of decision loss	19.3.3
Best expected loss	$\ell^*(y)$, eq. (19.15)	loss of the best action now	19.3.3
Active test	step of u_{test}	separates hypotheses by response	19.4.1
Safe test choice	eq. (19.18)	dual control under constraints	19.4.2
Fisher information	$\mathcal{I}(\theta) = J^{\mathsf{T}}\Sigma^{-1}J$	parameter information; D-optimal design	19.4.6
Sequential loop	Figure 19.8	measure, infer, choose, stop	19.5.2
Question-specific value	Table 19.4	one facility, several next-best measurements	19.5.4
Next-best model	residuals with low VOI everywhere	improve the model, not the data	19.6.1
Practical report	Table 19.6	hypotheses, measurement, reason, urgency	19.6.2

19.8 Exercises

Exercise 19.1. A diagnosis has posterior $(0.60, 0.30, 0.10)$ over three hypotheses. A candidate measurement has two outcome classes. Under the first, with predictive probability 0.6, the posterior would become $(0.95, 0.04, 0.01)$. (a) Use equation (19.9) to find the posterior after the second outcome. (b) Compute the entropy of the current posterior and of the two possible posteriors, in bits. (c) Compute the expected information gain and the percentage reduction of uncertainty. (d) Which hypotheses does this measurement separate, and which does it not?

Exercise 19.2. A pump is suspected of cavitation with probability 0.3. Repairing now costs \$15k whether or not the fault is present; waiting costs nothing if there is no fault and \$80k if there is. (a) Compute the expected loss of each action and the best action. (b) A vibration test costing \$2k resolves the question perfectly. Compute its value of information and its net value. (c) Find the fault probability at which the best action without the test switches, and the value of the test at that probability. (d) If both actions cost \$15k under the fault and nothing without it, what is the value of the test?

Exercise 19.3. Under H_1 a rack airflow reading is predicted as $\mathcal{N}(1.6, 0.25^2)\,\mathrm{kg/s}$; under H_2 as $\mathcal{N}(2.8, 0.3^2)\,\mathrm{kg/s}$; the two hypotheses have probabilities 0.5 each. (a) Compute the likelihood ratio and the posterior after a reading of $2.0\,\mathrm{kg/s}$. (b) A hot-aisle reading is predicted as $\mathcal{N}(37, 2^2)^\circ\mathrm{C}$ under H_1 and $\mathcal{N}(38, 2^2)^\circ\mathrm{C}$ under H_2. Compute the posterior after a reading of $37.5^\circ\mathrm{C}$. (c) Express the separation of the means of each measurement in standard deviations of their difference, and state which measurement should be taken and why.

Exercise 19.4. Two parameters $\theta = (\theta_1, \theta_2)$ have a prior information matrix equal to the identity. The existing sensor has sensitivity row $(1, 2)$ and noise standard deviation 1. (a) Compute the total Fisher information $\mathcal{I}$ and its determinant. (b) Compute both again after adding a second sensor identical to the first, with independent noise, and (c) after adding instead a sensor with row $(1, 0)$ and noise standard deviation 1. (d) By what factor does the area of the one-standard-deviation ellipse shrink in (b) and in (c), and what does the comparison say about diverse measurements?

19.9 Connection to the TERRA framework

Next-best measurement uses the inference the framework already performs and asks one further question of it: which additional measurement equation would change the posterior, and the decision, the most.

- **Hypotheses are alternative graphs or parameter sets.** H_1 is a reduced flow on the edge through the rack; H_2 is a recirculation edge from the hot-aisle node back to the rack-inlet node, carrying the fraction α; H_3 is a bias term b in the measurement equation of one inlet sensor. The framework's posterior over these alternatives is the $\Pr(H_i \mid y)$ of equation (19.1).
- **A candidate measurement is a new measurement equation attached to a node or an edge.** An airflow meter on the rack edge, a probe at the inlet node or a pressure sensor at the cold-aisle plenum is a function of the graph's states. Propagating the current posterior through that function under each hypothesis gives the predictive distributions of equation (19.2), and from them the information gain and the value of information.
- **An active test is a forward simulation under a perturbed control.** The fan-speed and topology tests of Figures 19.5 and 19.6 are simulations of the coupled air and thermal networks under a stepped fan input, run once per hypothesis; the component bounds

$T_{\text{in}} \le T_{\max}$ and $P_{\text{facility}} \le P^{\max}$ are checked on every run, which is constraint (19.18). Which unit serves which row is an uncertain edge of the air network that the test resolves.

- **Fisher information comes from the Jacobian the inference already assembles.** The sensitivity matrix J of equation (19.19) is the derivative of the measurement equations with respect to the inferred parameters such as $\dot{m}_{\text{air}}$ and α; the posterior covariance of a Laplace approximation is the inverse of the same information, and its ellipses are those of Figure 19.7.
- **Persistent residuals point to a missing edge.** When the balances at the aisle and rack nodes leave residuals that no hypothesis explains and every candidate has low value, the next step is a change to the graph, such as adding the recirculation edge, before any further measurement.

In summary, when the measurements cannot separate competing explanations, the next measurement is chosen by predicting its outcome under each hypothesis and valuing the expected change, in bits of diagnostic uncertainty or better in expected decision loss, against its cost and under safety constraints that hold for every plausible hypothesis; the choice is repeated after every result, depends on the question being asked, and gives way to a better model when the model itself is the limit. The next chapter turns from choosing what to learn to choosing what to do, in data center optimization (Chapter 20).

Chapter 20

Data Center Optimization

20.1 The optimization problem

The preceding chapters described the IT load, the power path, the cooling plant, the controls, the measurements and the ways in which the facility can fail, and they turned each of these into a model, a constraint or a probability. This chapter puts them together into one decision. Its question is

what should the data center do to minimize cost or energy while satisfying its thermal, electrical, reliability and computing constraints?

The answer is an optimization problem over the cooling plant, the IT load and the electrical resources together. The chapter states that problem, lists its constraints and objectives, treats workload, sites and storage as decisions, adds uncertainty in its robust, stochastic and risk-aware forms, adds the dynamic, discrete and contingency structure that makes the problem hard, and ends with the place of optimization in the operation of a facility: below the inference that supplies its inputs and above the low-level controls that carry out its decisions.

Jargon: objective, decision variable, constraint, feasible set, active constraint, dual value, robust and stochastic optimization, mixed-integer, unit commitment

The *objective* is the quantity an optimization minimizes, here a cost or an energy. The *decision variables* are the quantities the optimizer chooses, the control inputs u of the facility. A *constraint* is a condition the decision must satisfy, written $g(x, u) \leq 0$ so that a positive value is a violation, and the *feasible set* is the set of decisions that satisfy every constraint. A constraint is *active*, or *binding*, at a solution when it holds with equality there, so that relaxing it would improve the objective. The *dual value*, or *shadow price*, of a constraint is the rate at which the optimal objective would improve per unit of relaxation of that constraint. *Robust optimization* requires the constraints to hold for every value of an uncertain quantity in a declared set; *stochastic optimization* treats the uncertain quantities as random and optimizes an expectation, possibly with constraints on probabilities. A *mixed-integer* problem has some decision variables restricted to integer or on/off values. *Unit commitment* is the power-system problem of deciding which generating units to switch on over a planning period and how to load them; the same structure appears in the staging of chillers.

Key idea

Optimizing a data center is not minimizing PUE or electricity cost. It is choosing how to operate the IT, cooling and electrical resources together so that thermal safety, electrical capacity, service requirements and redundancy are respected; under uncertainty the question becomes which action gives the best performance while keeping the physical risk acceptably low.

This section states the generic problem, sorts the decision variables into three groups, and shows on the simplest objective, cooling energy, that the constraints rather than the objective carry the problem.

20.1.1 The generic problem

Every optimization in this chapter is a special case of

$$\min_{u} \; J(x,u) \qquad \text{subject to} \qquad g(x,u) \le 0, \qquad x_{k+1} = f(x_k, u_k, d_k, \theta), \tag{20.1}$$

with the symbols of Chapter 10: x is the state of the facility (temperatures, flows, powers, the state of charge of a battery, the work queued), u the decisions, d the disturbances (outdoor weather, the IT load that is not under control, equipment failures), θ the parameters (thermal resistances, coil conductances, equipment capacities and efficiencies), J the objective and g the vector of constraints. The third relation is the facility model in discrete time: the state at step $k+1$ follows from the state, the decision, the disturbance and the parameters at step k, through the balances of power, mass and energy developed in Chapters 3 to 8.

The form makes three facts explicit. The objective depends on the state as well as on the decision, because cooling energy depends on the heat the IT load produces. The constraints depend on the state too, because a rack temperature limit is a limit on x, which the optimizer reaches only through u and the model f. And the model contains d and θ, which are not known exactly; the first three sections of the chapter treat them as known, and Section 20.4 removes that assumption.

20.1.2 Decision variables

The decisions of a data center fall into three groups, listed in Table 20.1. Each row gives a decision, its symbol, whether it is continuous or discrete, and what it acts on. The cooling decisions u_{cool} are the actuator settings and setpoints of Chapter 10: the fan speed N_{fan} of the room air handlers, the pump speed N_{pump}, the coil valve opening u_{valve}, the chilled-water supply setpoint $T_{\text{CHW,set}}$ and the number of chillers running N_{ch}. The IT decisions u_{IT} are a power cap P_{cap} on servers or accelerators, the allocation of work to servers, the scheduling of jobs in time and the migration of workload between rooms or sites. The electrical decisions u_{elec} are the battery power P_{batt}, the state of the generators and the shedding of load. The full decision vector is

$$u = \left[\, u_{\text{cool}},\; u_{\text{IT}},\; u_{\text{elec}} \,\right]. \tag{20.2}$$

Figure 20.1 draws how the three groups enter the problem. The three boxes on the left are the decision groups. The large box in the middle is the facility model f, which couples three conservation networks (electrical power, air and water mass, thermal energy); the disturbances and parameters enter it from above. The two boxes on the right are what the optimizer evaluates for every candidate decision: the constraints and the objective. Each decision group acts on more than one network. A power cap reduces electrical power and, because nearly all IT power becomes heat, the heat the cooling must remove; a higher chilled-water setpoint raises the chiller's efficiency and the temperature of the air reaching the racks. The coupling is why the groups are optimized together rather than one at a time.

Table 20.1: The decision variables of a data center, by group.

Group	Decision	Symbol	Type	Acts on
cooling	air handler fan speed	N_{fan}	continuous	room airflow, rack inlet temperatures
	chilled-water pump speed	N_{pump}	continuous	loop flow and pressure
	coil valve opening	u_{valve}	continuous	water flow through one coil
	chilled-water supply setpoint	$T_{\text{CHW,set}}$	continuous	chiller efficiency, supply air temperature
	chillers running	N_{ch}	integer	plant capacity and margin
IT	power cap	P_{cap}	continuous	IT power and heat, computing speed
	server allocation, job scheduling	–	discrete	where and when heat is produced
	workload migration	–	discrete	which room or site carries the load
electrical	battery power	P_{batt}	continuous	grid draw, state of charge
	generator state	–	on/off	supply source, fuel
	load shedding	–	on/off	served load

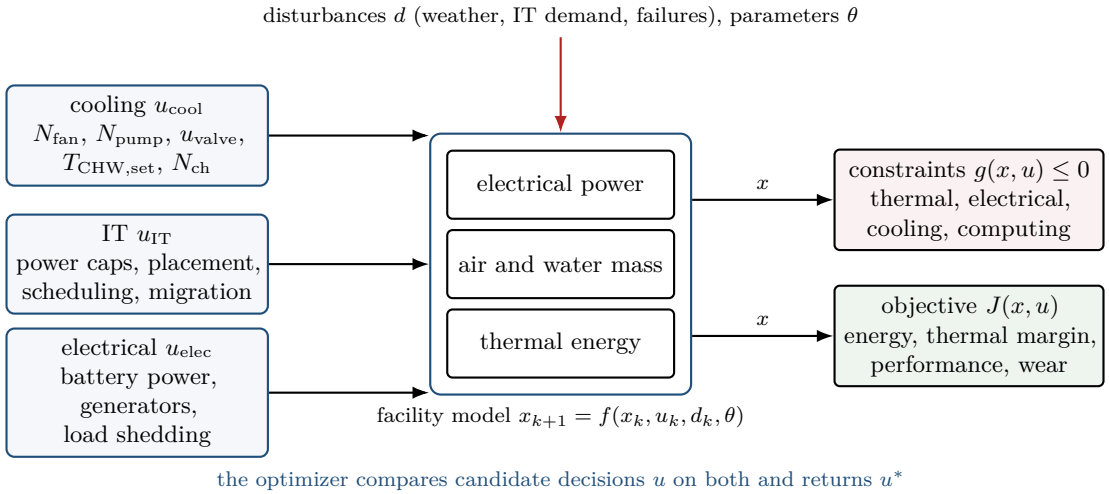


Figure 20.1: The data center optimization problem. Three groups of decisions act through the facility model, which couples the electrical, mass and thermal networks under disturbances and uncertain parameters; every candidate decision is checked against the constraints and scored by the objective.

20.1.3 The cooling-energy objective

The simplest useful problem minimizes the electrical power of the cooling system subject to a temperature limit at every rack,

$$\min_{u_{\text{cool}}} P_{\text{cool}} \qquad \text{subject to} \qquad T_{\text{in},i} \le T_{\max} \quad \text{for every rack } i, \tag{20.3}$$

where $T_{\text{in},i}$ is the inlet temperature of rack i. The direction of the solution is known before any calculation. Lowering the fan speed reduces fan power by the cube law of Chapter 6, and raising the chilled-water setpoint reduces chiller power because the chiller lifts heat across a smaller temperature difference (Chapter 7). Both moves also raise rack inlet temperatures. The optimizer therefore lowers N_{fan} and raises $T_{\text{CHW,set}}$ until a thermal constraint becomes active, and the solution lies on the boundary of the feasible set, not inside it.

An illustrative two-variable version makes this concrete. Express the fan speed N_{fan} as a fraction of its rated speed and take the cooling power in megawatts as

$$P_{\text{cool}} = 0.6\, N_{\text{fan}}^3 + 1.6\,[\,1 - 0.04\,(T_{\text{CHW,set}} - 7)\,], \tag{20.4}$$

a fan system of 0.6 MW at rated speed and a chiller plant of 1.6 MW at a 7°C setpoint whose power falls by 4 % per kelvin of setpoint. Take the inlet temperature of the hottest rack as

$$T_{\text{in,hot}} = T_{\text{CHW,set}} + 14 + \frac{8}{N_{\text{fan}}} \ \ ^\circ\text{C}, \tag{20.5}$$

an illustrative fit that lumps the coil approach, the heating of the air on its way to the rack and a recirculation that grows as the airflow falls. The limit is $T_{\max} = 35$°C with a 2 K margin, so the constraint is $T_{\text{in,hot}} \le 33$°C, that is $T_{\text{CHW,set}} \le 19 - 8/N_{\text{fan}}$. On that boundary the cooling power is a function of the fan speed alone, $P_{\text{cool}} = 0.6\, N_{\text{fan}}^3 + 0.832 + 0.512/N_{\text{fan}}$, whose derivative vanishes at $N_{\text{fan}}^4 = 0.512/1.8$, that is $N_{\text{fan}} = 0.730$, with $T_{\text{CHW,set}} = 8.05$°C and $P_{\text{cool}} = 1.767$ MW. The conventional operating point, full fan speed and a 7°C setpoint, draws 2.2 MW with the hottest inlet at 29°C; the optimum saves 0.43 MW, about 20 %, by spending the 4 K of slack that the conventional point leaves unused.

Figure 20.2 draws the problem. The horizontal axis is the chilled-water setpoint and the vertical axis the fan speed as a fraction of rated speed. The shaded region is infeasible: there the hottest inlet exceeds 33°C. The thick boundary is the active constraint. The thin curves are contours of constant cooling power from equation (20.4), labeled in megawatts. Cooling power falls toward the lower right, where the fan is slow and the water warm, and that is exactly where the constraint forbids the optimizer to go. The optimum is the point at which a contour just touches the boundary: the 1.767 MW contour is tangent to it at (8.05°C, 0.73), and the 1.6 MW contour lies wholly in the infeasible region. The figure also shows that neither variable is pushed to its extreme: lowering the fan further would require a colder setpoint to hold the hottest inlet, and the chiller power that costs would exceed the fan power saved.

The unconstrained optimum. Without the temperature constraint, equation (20.3) has a trivial solution: switch off the fans and the chillers, and $P_{\text{cool}} = 0$. The problem is not interesting because of its objective, which any operator can minimize by turning equipment off; it is interesting because of its constraints, which say how far the equipment can be turned down before something overheats, trips or fails to serve its load. Most of the work of data center optimization is therefore the work of writing the constraints correctly and knowing the state well enough to know where they bind.

> **Principle 20.1.** *In data center optimization the constraints carry the problem. The objective points toward less cooling, less margin and less redundancy; the thermal, electrical, cooling and computing constraints say how far that direction may be followed, and the solution lies where they become active.*

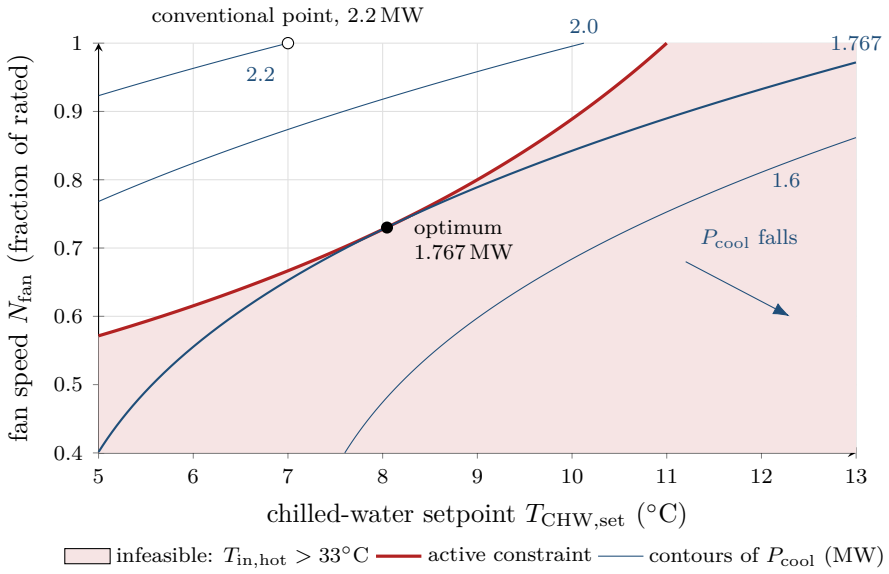


Figure 20.2: The cooling-energy problem of equation (20.3) in two variables. Contours of $P_{\text{cool}} = 0.6\,N_{\text{fan}}^3 + 1.6[1 - 0.04(T_{\text{CHW,set}} - 7)]$ in MW; infeasible region $T_{\text{CHW,set}} + 14 + 8/N_{\text{fan}} > 33$°C (shaded). The optimum lies on the active constraint at $N_{\text{fan}} = 0.73$, $T_{\text{CHW,set}} = 8.05$°C. Illustrative values.

20.2 Constraints and objectives

This section writes the four families of constraints, thermal, electrical, cooling and computing, and then replaces the single cooling-power objective with a weighted objective whose main term is the cost of energy.

20.2.1 Thermal constraints

The first family limits temperatures. At the rack level the limit is on inlet air,

$$T_{\text{in},i} \le T_{\max} \quad \text{for every rack } i, \qquad T_{\text{CPU},j} \le T_{\text{CPU,max}} \quad \text{for every processor } j, \tag{20.6}$$

where the second inequality limits the temperature of a component, here a processor, reported by the server's own telemetry. The two act on different time scales: an inlet limit protects the equipment class as a whole and is the one the facility controls directly, while a component limit is enforced by the server, which throttles its clock when the component overheats and so converts a thermal violation into a loss of computing speed.

An optimizer that drives the facility exactly to $T_{\max}$ has no room for anything it does not know. The constraint is therefore usually written with a margin,

$$T_{\text{in},i} \le T_{\max} - \Delta T_{\text{margin}}, \tag{20.7}$$

as in Section 20.1.3, where a 2 K margin under a 35°C limit gave the 33°C boundary. The margin absorbs model error, sensor bias, the transient after a load step and the racks whose inlets are not measured. It is a crude form of a chance constraint: a fixed number of kelvin standing in for a quantile of the uncertainty, and Section 20.4.3 replaces it with that quantile. Of the quantities in equation (20.6), an inlet temperature is measured at a few points per rack, a processor temperature is measured by the server, and the worst inlet in a row of unmeasured racks is inferred.

20.2.2 Electrical constraints in normal and contingency states

The second family limits power. The facility draw must stay within the capacity of its utility connection, and each uninterruptible power supply (UPS) within its rating,

$$P_{\text{facility}} \le P_{\text{grid,max}}, \qquad P_{\text{UPS},j} \le P_{\text{UPS,max},j}. \tag{20.8}$$

These hold in the normal state. A redundant power path exists so that the facility survives the loss of a component, so the constraints must also hold after each credible contingency c, a defined failure such as the loss of one UPS:

$$P^{(c)}_{\text{UPS},j} \le P_{\text{UPS,max},j} \quad \text{for every surviving UPS } j \text{ and every contingency } c, \tag{20.9}$$

where the superscript (c) marks the value in the post-contingency state. The problem therefore contains several copies of the facility: the normal one and one for each contingency, as in contingency analysis of a transmission system.

A $2N$ power path shows what the post-contingency constraint does. Two UPS systems A and B, each rated 1.0 MW, share a critical load. With 1.6 MW of load each carries 0.8 MW, 80 % of its rating, and equation (20.8) is satisfied. After the loss of A, B carries the whole 1.6 MW, which violates equation (20.9) by 0.6 MW. The post-contingency constraint limits the critical load to 1.0 MW, each UPS to 50 % of its rating in the normal state, which is the familiar loading rule of a $2N$ design (Chapter 4) derived rather than assumed. An optimizer that sees only the normal state would fill the path to 2.0 MW and leave it without redundancy.

20.2.3 Cooling-equipment and computing constraints

The third family bounds the cooling equipment:

$$\dot{Q}_{\text{ch},i} \le \dot{Q}^{\max}_{\text{ch},i}, \qquad \dot{m}_w \le \dot{m}^{\max}_w, \qquad N_{\text{fan,min}} \le N_{\text{fan}} \le N_{\text{fan,max}}, \tag{20.10}$$

for the heat $\dot{Q}_{\text{ch},i}$ removed by chiller i, the water flow through the loop and the fan speed. The upper bounds are capacities; the lower bound on fan speed has its own reasons. Below a minimum speed the air handlers cannot hold the plenum pressure that distributes air to the far racks, a motor may run outside its stable range, and the recirculation of equation (20.5) grows faster than the fan power falls. A chiller has a similar minimum stable loading, which Section 20.5.3 uses.

The fourth family protects the purpose of the facility. The work served must meet the work required, and the response time of a service must stay within its limit:

$$W_{\text{served}} \ge W_{\text{required}}, \qquad t_{\text{lat}} \le t_{\text{lat,max}}, \tag{20.11}$$

where W is computing work in whatever unit the service counts (jobs, requests, accelerator hours) and t_{lat} is the network and processing latency seen by a user. Without these constraints the cheapest data center is an idle one: shedding the IT load removes the IT power and the cooling power with it. With them, workload can be moved in time or in space, slowed within its deadline or capped within its latency, but not cut arbitrarily to save energy. Table 20.2 collects the four families with an example of each and the way the constrained quantity is usually known.

20.2.4 The weighted objective and the cost of energy

Cooling power alone is too narrow an objective. A more realistic one weighs four costs,

$$J = w_1\, C_{\text{energy}} + w_2\, C_{\text{thermal}} + w_3\, C_{\text{perf}} + w_4\, C_{\text{wear}}, \tag{20.12}$$

Table 20.2: The four families of constraints, with illustrative limits and whether the constrained quantity is measured, calculated or inferred.

Family	Constraint	Example	How the quantity is known
thermal	$T_{\text{in},i} \le T_{\max} - \Delta T_{\text{margin}}$	35°C less 2 K	measured at a few racks; inferred at the others
electrical	$P_{\text{facility}} \le P_{\text{grid,max}}$; $P^{(c)}_{\text{UPS},j} \le P_{\text{UPS,max},j}$	100 MW; 1.0 MW per UPS	facility power measured; post-contingency loading calculated
cooling	$\dot{Q}_{\text{ch},i} \le \dot{Q}^{\max}_{\text{ch},i}$; $N_{\text{fan,min}} \le N_{\text{fan}}$	2.5 MW per chiller; 40 % fan speed	chiller duty calculated; available capacity of a degraded unit inferred
computing	$W_{\text{served}} \ge W_{\text{required}}$; $t_{\text{lat}} \le t_{\text{lat,max}}$	job deadline 08:00; 50 ms	work counted by the scheduler; latency measured

where C_{energy} is the cost of the electricity drawn, C_{thermal} a penalty for operating close to thermal limits (a cost of lost margin, short of a violation), C_{perf} the cost of slower, deferred or migrated computing, and C_{wear} the cost of starts, cycling and running hours of equipment. The weights w_1 to w_4 convert the four terms into one currency. Choosing them is a management decision about how much energy is worth a unit of thermal margin or an hour of delayed work, and the merit of writing them down is that the exchange rates become visible and can be reviewed.

The energy term is a sum over time steps of length Δt,

$$C_{\text{energy}} = \sum_{t} p_t \, P_{\text{facility},t} \, \Delta t, \tag{20.13}$$

with p_t the price of electricity in step t in dollars per megawatt hour. For a facility drawing a constant 30 MW through a day priced at an illustrative \$40 per MWh for 8 night hours, \$60 for 10 day hours and \$120 for 6 peak hours, the cost is $30 \times (8 \times 40 + 10 \times 60 + 6 \times 120) = 30 \times 1640 =$ \$49,200. The peak hours are a quarter of the day and 44 % of the cost. Once the price varies in time, the time at which energy is used becomes a decision variable in its own right, which the next section develops.

20.3 Workload, sites and storage as decisions

This section treats the IT load and the battery as decisions rather than as givens: computing work moved in time against a price, moved between sites against the weather and the latency limit, traded against cooling effort on a hot afternoon, and electrical energy moved in time by a battery whose backup reserve has a price of its own.

20.3.1 Electricity price and flexible load

With a time-varying price $p(t)$ the same computation costs different amounts at different times. With the illustrative prices of the next subsection, a megawatt hour drawn at 02:00 costs \$30 and one drawn at 17:00 costs \$120, four times as much. Work that has no deadline in the next hours, such as model training, batch analytics, backups or rendering, can therefore be shifted from expensive hours to cheap ones, $W(t)$ moving to where $p(t)$ is low, provided its deadline allows. This is flexible-load scheduling. Interactive services cannot be shifted in time, because a

user's request must be served when it arrives; the flexible share of a facility's load is a property of its workload mix and is agreed with the owners of that workload.

The incentive is larger than the IT power alone. Every megawatt of IT load moved also moves the cooling and distribution losses it brings with it, and these are usually smaller at night, when the outdoor air is cool and the chillers lift heat across a smaller temperature difference. The incremental facility power of a job is therefore its IT power multiplied by an incremental overhead factor, which is higher in the hot afternoon than at night. Shifting a job moves it to a cheaper price and to a smaller overhead at the same time.

20.3.2 A batch-job schedule

A batch job must receive $E_{\mathrm{job}} = 10\,\mathrm{MWh}$ of IT energy before 08:00. It arrives at 16:00 and can run at any power between zero and $P_{\mathrm{job}}^{\max} = 2.5\,\mathrm{MW}$, so it needs at least four full hours in the sixteen available. The schedule $P_{\mathrm{job}}(t)$ is the solution of

$$\min_{P_{\mathrm{job}}} \sum_t p_t\,\kappa_t\,P_{\mathrm{job},t}\,\Delta t \quad \text{subject to} \quad \sum_t P_{\mathrm{job},t}\,\Delta t = E_{\mathrm{job}}, \quad 0 \le P_{\mathrm{job},t} \le P_{\mathrm{job}}^{\max}, \tag{20.14}$$

with the sum running over the hours from 16:00 to 07:00 and κ_t the incremental overhead factor, taken as 1.4 from 16:00 to 19:59 and 1.2 at other hours (illustrative). The continuous form of the energy constraint is $\int P_{\mathrm{job}}\,\mathrm{d}t = 10\,\mathrm{MWh}$. Because the objective and the constraints are linear and there is a single energy constraint, the solution fills the hours with the lowest $p_t\kappa_t$ to the power limit, one after another, until the energy is delivered.

The illustrative hourly prices, in dollars per megawatt hour, are 80, 120, 140 and 110 for the four hours from 16:00, then 85, 70, 60, 50, 42, 36, 30, 28, 30, 35, 48 and 65 for the hours from 20:00 to 07:00. A scheduler that starts the job on arrival runs it from 16:00 to 19:59 and pays

$$2.5 \times 1.4 \times (80 + 120 + 140 + 110) = 3.5 \times 450 = \$1575. \tag{20.15}$$

The optimal schedule runs it in the four cheapest hours, 02:00 to 05:59 at \$30, \$28, \$30 and \$35:

$$2.5 \times 1.2 \times (30 + 28 + 30 + 35) = 3.0 \times 123 = \$369. \tag{20.16}$$

The saving is \$1206, 77 % of the cost on arrival. The optimization has a second effect, which the price alone does not show: the on-arrival schedule adds 2.5 MW of heat at 18:00, the hour in which the cooling plant has the least margin on a hot day, and the optimal schedule removes it. A constraint $P_{\mathrm{job},18} = 0$ imposed by the cooling plant would not change the optimal schedule at all, while it would make the on-arrival schedule infeasible.

Figure 20.3 draws the two schedules. The upper panel is the price against the time of day as a step function; the lower panel is the job power in megawatts on the same time axis, gray for the on-arrival schedule and green for the optimal one, with the 18:00 hour of least cooling margin marked in red. The figure shows the schedule following the bottom of the price curve and the on-arrival schedule sitting on its peak. The deadline, the energy and the power limit determine how wide a band of cheap hours the job needs; a power limit of 5 MW would let it run in the two cheapest hours alone.

20.3.3 Geographic shifting and latency

An operator with more than one site can move work in space as well as in time. Suppose site A sits under a 38°C afternoon and site B under 18°C. At B the cooling plant can reject heat to the outdoor air with little or no compressor work, while at A the chillers lift heat across a large temperature difference. With illustrative incremental overhead factors of 1.45 at A and 1.12 at B, moving 4 MW of IT load from A to B reduces the facility draw at A by $4 \times 1.45 = 5.8\,\mathrm{MW}$ and raises it at B by $4 \times 1.12 = 4.48\,\mathrm{MW}$, a net saving of 1.32 MW; over a six-hour afternoon

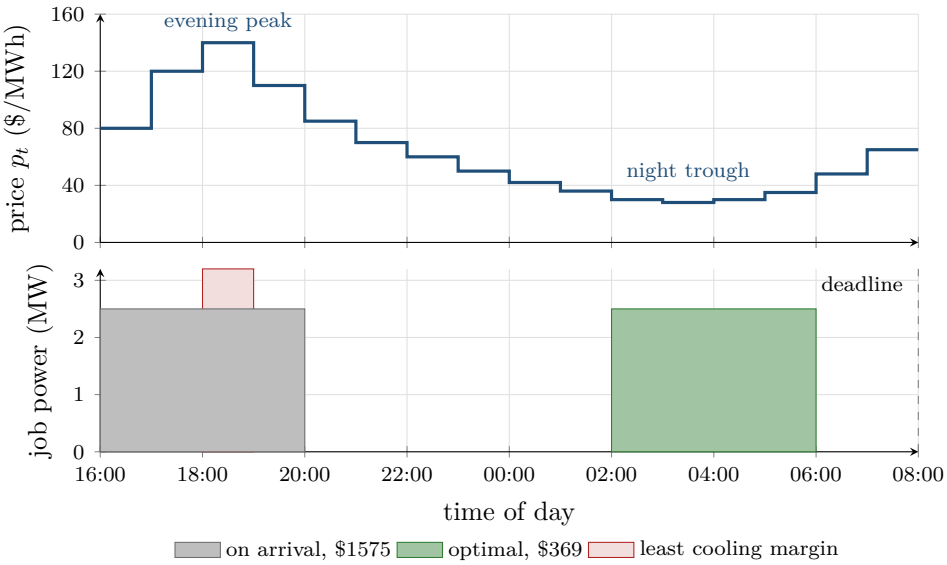


Figure 20.3: A 10 MWh batch job scheduled against the hourly price. Upper panel: illustrative prices. Lower panel: the on-arrival schedule (16:00 to 19:59) and the solution of equation (20.14) (02:00 to 05:59), both at $P_{\text{job}}^{\max} = 2.5$ MW, with incremental overhead factors 1.4 in the evening and 1.2 otherwise. Illustrative values.

at \$80 per MWh that is \$634. The decision variables are now the workloads W_A and W_B, with W_A lowered and W_B raised, and the problem is an optimization across sites with a shared constraint that the total work is served.

The shift is limited by the computing constraint of equation (20.11). If the users of a service are near site A, serving them from B adds network delay. With illustrative values, an interactive service whose limit is $t_{\text{lat,max}} = 50$ ms, whose users see 20 ms from A and whose path to B adds 35 ms, cannot be moved: 55 ms exceeds the limit. A training job whose users are not waiting for each response can be moved freely. The geographic problem is therefore not purely an energy problem: it is an energy problem constrained by where the users are, by the data a job needs and by the capacity at the receiving site, which must itself satisfy every constraint of Section 20.2.

20.3.4 Joint cooling and workload optimization

Traditional facility control treats $P_{\text{IT}}(t)$ as a disturbance: the IT load is what it is, and the cooling is optimized to carry it. When part of the IT load is flexible it becomes a decision, and the cooling decisions u_{cool} and the IT decisions u_{IT} are optimized together, as in the joint problem of Chapter 10. The gain comes from the coupling noted in Section 20.1.2: reducing the heat at its source can be cheaper than removing it.

Consider a hot afternoon, 14:00 to 17:00, on which the running chillers can remove 6.0 MW and the forecast cooling requirement is 6.3 MW, a shortfall of 0.3 MW; the forecast has a standard deviation of 0.2 MW and the price is \$150 per MWh. Two options are on the table. In option A the plant cools harder: it starts another chiller and raises fan and pump speeds, which adds 1 MW of cooling power for the three hours and 2.0 MW of cooling capacity. In option B the facility defers 0.5 MW of noncritical compute to the night, which lowers the heat load by 0.5 MW and possibly avoids the chiller start. Table 20.3 compares them. Each row gives the change in facility power, the cost items, the margin of available over required cooling, and the probability that the requirement exceeds the capacity under a normal forecast. Option A costs $3 \times 1 \times 150 = \$450$ of energy and an illustrative \$300 for the start and its wear, \$750 in total,

and leaves a margin of 1.7 MW. Option B saves $0.5 \times 1.2 \times 3 = 1.8$ MWh in the afternoon, worth \$270, spends the same energy at night at \$40 per MWh, \$72, and incurs a performance cost of \$200 per MWh of deferred work on 1.5 MWh, \$300: a net \$102.

Table 20.3: Two responses to a hot afternoon with a 0.3 MW forecast cooling shortfall over three hours; forecast standard deviation 0.2 MW. Illustrative values.

Option	Facility power	Cost	Margin	$\Pr(\dot{Q}_{\text{req}} > \dot{Q}_{\text{avail}})$
no action	–	–	−0.3 MW	0.93
A: cool harder (start a chiller)	+1.0 MW	\$450 energy + \$300 start = \$750	+1.7 MW	≈ 0
B: defer 0.5 MW of compute	−0.6 MW	−\$270 + \$72 + \$300 = \$102	+0.2 MW	0.16

On cost, option B wins by \$648. On risk it loses: its 0.2 MW margin is one forecast standard deviation, and the probability that the requirement still exceeds the capacity is $1 - \Phi(1) = 0.16$, where Φ is the standard normal distribution function. Which option is better therefore depends on three quantities the table makes explicit: the computing penalty, which would have to exceed $(750 + 198)/1.5 = \$632$ per MWh of deferred work before A became the cheaper option; the electricity price, which sets both options' energy terms; and the thermal margin and its uncertainty. If the forecast standard deviation were 0.08 MW instead of 0.2 MW, option B would leave a probability of $1 - \Phi(2.5) = 0.006$ and would be both cheaper and acceptably safe. A joint optimizer evaluates exactly this trade, and a controller that treats the IT load as fixed never sees option B.

20.3.5 The battery and its reserve constraint

A battery, in the UPS or as a separate storage system, is the electrical decision variable. With $P_{\text{batt}} > 0$ on discharge, it can reduce the facility's peak draw, keep it under the utility capacity limit, hold a backup reserve, and offer services to the grid (a later chapter). Its state of charge evolves as

$$\text{SOC}_{t+1} = \text{SOC}_t - \frac{P_{\text{batt},t}\,\Delta t}{E_{\text{usable}}}, \tag{20.17}$$

with E_{usable} the usable energy; conversion losses are omitted. With $E_{\text{usable}} = 20$ MWh, a 5 MW discharge for one hour lowers the state of charge by 0.25.

The backup role is a constraint,

$$\text{SOC}_t \geq \text{SOC}_{\text{reserve}}, \tag{20.18}$$

for example $\text{SOC}_{\text{reserve}} = 0.8$, which keeps 16 MWh for ride-through: at an illustrative critical load of 80 MW, twelve minutes for the generators to start and take the load. Suppose the facility will peak at 103 MW for two hours against a utility limit of $P_{\text{grid,max}} = 100$ MW. Shaving the peak needs $3 \times 2 = 6$ MWh from a battery that starts the day full, but the reserve allows only $(1 - 0.8) \times 20 = 4$ MWh. The remaining 2 MWh must come from deferring flexible compute, at an illustrative penalty of \$300 per MWh, so the optimal cost of the afternoon is \$600. The reserve constraint is active.

Figure 20.4 plots the optimal cost J^* of that afternoon against the reserve level. The horizontal axis is $\text{SOC}_{\text{reserve}}$ and the vertical axis J^* in dollars; the labels beside the marked points give the ride-through time the reserve provides at 80 MW. Below a reserve of 0.7 the battery covers the whole 6 MWh and the cost is zero; above it every unit of reserve displaces battery energy that must be replaced by deferred compute. The slope of the line is the dual

value of the reserve constraint: 20 MWh × \$300 per MWh = \$6000 per unit of state of charge, \$60 per percentage point. Raising the reserve from 0.8 to 0.9 costs \$600 that afternoon and buys 1.5 minutes of ride-through; lowering it to 0.7 saves \$600 and gives up 1.5 minutes. The optimizer computes the price of resilience; whether it is worth paying is a decision about the consequence of a failed ride-through, which Section 20.4.4 takes up.

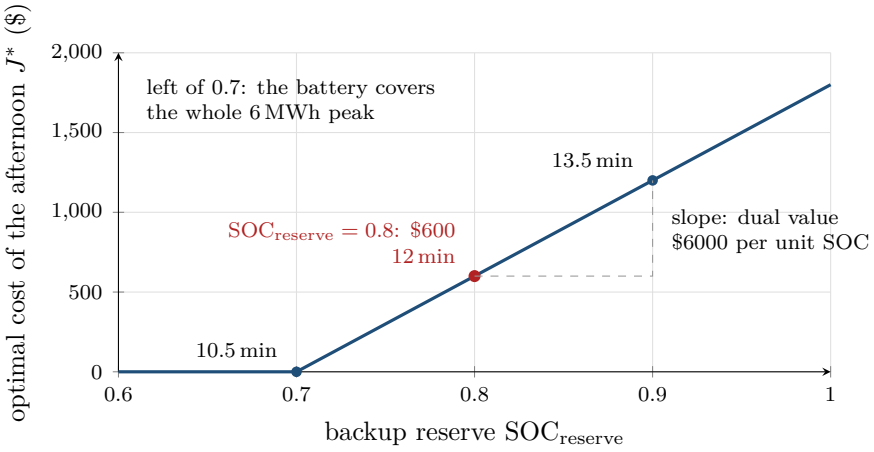


Figure 20.4: The price of the backup reserve. Optimal cost $J^* = 300\max(0,\ 6 - 20(1 - \text{SOC}_{\text{reserve}}))$ dollars of shaving a 6 MWh peak with a 20 MWh battery, with deferred compute at \$300 per MWh covering what the battery cannot. Labels: ride-through at 80 MW, 15 $\text{SOC}_{\text{reserve}}$ minutes. Illustrative values.

20.4 Optimization under uncertainty

Every problem so far assumed that the IT load, the weather, the equipment capacities and the thermal parameters were known. This section removes the assumption and develops the three standard answers, robust optimization, stochastic optimization and chance constraints, then sets risk tolerances by consequence and adds a measure of tail risk to the objective.

20.4.1 Deterministic inputs and real uncertainty

A deterministic optimization takes P_{IT}, T_{amb}, $\dot{Q}_{\text{avail}}$ and R_{th} as known and solves

$$\min_{u}\ J(x,u) \quad \text{subject to} \quad g(x,u) \le 0, \tag{20.19}$$

with every constraint evaluated at those values. Its answer is the best decision if the inputs are right, and it sits on the boundary of the feasible set, as Figure 20.2 showed. A decision on a boundary computed from wrong inputs is on the wrong side of the true boundary half the time.

In operation the inputs are not known. The IT load $P_{\text{IT}}(t+h)$ at a horizon h ahead is a forecast. The weather is a forecast. The capacity of a chiller with a fouled condenser is below its nameplate by an amount that must be estimated. The airflow through a rack is rarely measured and is inferred from temperatures and fan speeds, and every temperature carries a possible sensor bias b. The honest description of the inputs is a set of distributions,

$$d \sim p(d), \qquad \theta \sim p(\theta), \qquad x \sim p(x \mid y), \tag{20.20}$$

a forecast distribution for the disturbances, a distribution for the parameters and a posterior for the state given the measurements y, as developed in Chapters 14 and 15. The three subsections that follow differ in what they do with these distributions.

20.4.2 Robust optimization and its conservatism

Robust optimization declares a set Θ of plausible parameter values and requires every constraint to hold for all of them,

$$g(x, u, \theta) \le 0 \quad \text{for all } \theta \in \Theta, \tag{20.21}$$

for example $T_{\text{in}} \le 35°\text{C}$ for every plausible airflow. It protects against the worst case in the set and is safe by construction, provided the set contains the truth.

Take an illustrative rack whose servers draw $\dot{m}_s = 2.0\,\text{kg/s}$ of air and heat it by $\Delta T = 20\,\text{K}$ (a 40 kW rack). The cold aisle delivers $\dot{m}$ to the rack; when $\dot{m} < \dot{m}_s$, the shortfall is drawn from the hot exhaust, so the recirculation fraction is $\alpha = 1 - \dot{m}/\dot{m}_s$. The mixing rule of Chapter 6, $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{out}}$ with $T_{\text{out}} = T_{\text{in}} + \Delta T$, gives

$$T_{\text{in}} = T_{\text{supply}} + \Delta T\,\frac{\alpha}{1-\alpha} = T_{\text{supply}} + 20\,\frac{2.0 - \dot{m}}{\dot{m}} \quad °\text{C}. \tag{20.22}$$

The delivered airflow is inferred, with an estimate of 1.5 kg/s and a standard deviation of 0.105 kg/s under a normal model. At 1.5 kg/s the inlet sits 6.7 K above the supply air, so the highest supply temperature satisfying $T_{\text{in}} \le 35°\text{C}$ is 28.3°C. A robust optimizer whose set includes an airflow as low as 1.0 kg/s must hold the inlet below 35°C there too, where it sits 20 K above the supply air, and so restricts the supply air to 15.0°C.

Under the normal model, an airflow of 1.0 kg/s lies $0.5/0.105 = 4.75$ standard deviations below the estimate, a probability of about 10^{-6}. The worst-case controller acts as if that case were always possible and holds the supply air 13.3 K colder than the estimate requires; at the 4 % per kelvin of equation (20.4) that is more than half of the chiller power, spent every hour to protect against an event expected once in a million hours. Robust optimization is too conservative when the set is drawn wide enough to be safe, and unsafe when it is drawn narrow enough to be economical: it uses the range of the uncertainty and discards its distribution.

20.4.3 Stochastic optimization and chance constraints

Stochastic optimization uses the distribution. Its objective is an expectation and its constraints are probabilities,

$$\min_{u}\ \mathbb{E}\big[J(x, u, d, \theta)\big] \quad \text{subject to} \quad \Pr\big(T_{\text{in}} > 35°\text{C}\big) < 0.01. \tag{20.23}$$

The general chance constraint, introduced in Chapter 10, is

$$\Pr\big(g(x, u) \le 0\big) \ge 1 - \epsilon, \tag{20.24}$$

with ϵ the risk tolerance: for example $\Pr(T_{\text{in}} \le 35°\text{C}) \ge 0.99$ and $\Pr(P_{\text{facility}} \le 100\,\text{MW}) \ge 0.999$. For a normal prediction of a temperature with mean μ and standard deviation σ, the limit $\Pr(T > T_{\max}) \le \epsilon$ is equivalent to $\mu + z_{1-\epsilon}\sigma \le T_{\max}$, with $z_{1-\epsilon}$ the standard normal quantile, and the same holds for a power and its limit; the term $z_{1-\epsilon}\sigma$ is the back-off from the limit. Two worked cases follow.

For a facility power forecast with $\sigma = 1.5\,\text{MW}$ and $\epsilon_E = 0.001$, $z_{0.999} = 3.090$ and the planned mean draw must satisfy $\mu \le 100 - 3.090 \times 1.5 = 95.4\,\text{MW}$: the back-off is 4.6 MW. For a rack inlet prediction with $\sigma = 1.2\,\text{K}$ and $\epsilon_T = 0.01$, $z_{0.99} = 2.326$ and the planned mean must satisfy $\mu \le 35 - 2.326 \times 1.2 = 32.2°\text{C}$: a back-off of 2.8 K, which replaces the fixed 2 K margin of equation (20.7) with one computed from the uncertainty.

For the rack of equation (20.22) the uncertainty sits in the airflow, and the inlet temperature is a decreasing function of it. The constraint $\Pr(T_{\text{in}} > 35°\text{C}) \le 0.01$ then holds exactly when the inlet stays below 35°C at the first percentile of the airflow, $1.5 - 2.326 \times 0.105 = 1.255\,\text{kg/s}$. There the inlet is $20 \times 0.745/1.255 = 11.9\,\text{K}$ above the supply air, and the highest allowed supply temperature is 23.1°C: 5.2 K below the value at the estimate, 8.1 K above the robust value.

Figure 20.5 draws the whole range. The horizontal axis is the risk tolerance ϵ on a logarithmic scale and the vertical axis the highest supply air temperature that satisfies $\Pr(T_{\text{in}} > 35°\text{C}) \le \epsilon$. The solid curve is the airflow estimate of this subsection, $\sigma = 0.105\,\text{kg/s}$; the dashed curve is the same rack with a better airflow estimate, $\sigma = 0.05\,\text{kg/s}$, which Section 20.6.4 uses. The horizontal reference lines are the value at the estimate, reached at $\epsilon = 0.5$, and the robust value. Reading the solid curve from right to left: each tenfold reduction in ϵ costs between 2.0 and 2.6 K of supply temperature, and the curve meets the robust line at 10^{-6}, the probability of the robust worst case. The dashed curve lies above the solid one everywhere: a narrower posterior buys back supply temperature at every tolerance, 3.0 K at $\epsilon = 0.01$.

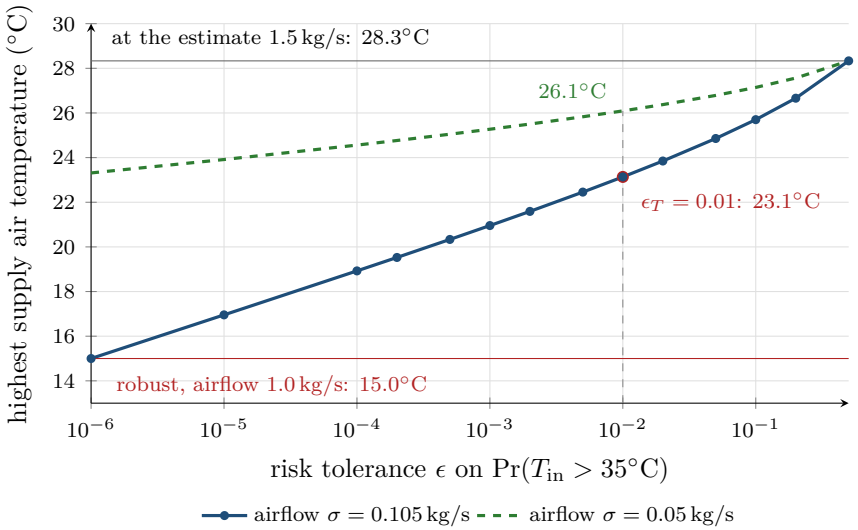


Figure 20.5: The price of conservatism for the rack of equation (20.22): the highest supply temperature $35 - 20(2 - \dot{m}^*)/\dot{m}^*$ with $\dot{m}^* = 1.5 - z_{1-\epsilon}\sigma$ the ϵ-quantile of a normal airflow estimate, for $\sigma = 0.105\,\text{kg/s}$ (solid) and $0.05\,\text{kg/s}$ (dashed). The solid curve reaches the robust value at $\epsilon \approx 10^{-6}$. Illustrative values.

Principle 20.2. *Uncertainty enters an optimization as a back-off from each limit. A robust formulation sets the back-off by the edge of a plausible set and is usually too conservative; a chance constraint sets it by a quantile, $z_{1-\epsilon}\sigma$ for a normal prediction, so that it grows with the uncertainty of the estimate and shrinks when the estimate improves.*

20.4.4 Risk tolerance by consequence

The tolerance ϵ is not a single number for the whole facility. It is set by the consequence of the violation it limits, and Table 20.4 lists three illustrative levels. Each row gives the event, its tolerance, the normal quantile that sets the back-off, and the back-off for an illustrative uncertainty. A rack inlet briefly above its limit costs a little equipment life and possibly some throttling, and is tolerated at $\epsilon_T = 0.01$. An electrical overload trips a breaker or overloads a utility connection and is tolerated at $\epsilon_E = 0.001$. An interruption of a critical service is tolerated at $\epsilon_S = 10^{-5}$. The more severe the consequence, the smaller the tolerance, and the larger the back-off for the same uncertainty: the quantile grows from 2.33 to 4.26 standard deviations across the table.

Setting tolerances by consequence is what lets a facility operate close to its limits where a violation is cheap and far from them where it is not. A single tolerance for every constraint

Table 20.4: Risk tolerances set by consequence, with the back-off each implies for a normal prediction. Illustrative values.

Event	Tolerance	$z_{1-\epsilon}$	Back-off for an illustrative σ
rack inlet above 35°C	$\epsilon_T = 0.01$	2.326	$\sigma = 1.2$ K: 2.8 K
facility power above 100 MW	$\epsilon_E = 0.001$	3.090	$\sigma = 1.5$ MW: 4.6 MW
critical-service interruption	$\epsilon_S = 10^{-5}$	4.265	$\sigma = 1.5$ MW of ride-through demand: 6.4 MW

would either waste energy on the minor constraints or accept too much risk on the major ones. The tolerances are inputs to the optimizer, chosen by the people who answer for the consequences, in the same way as the weights of equation (20.12).

20.4.5 Expected cost and tail risk

An expected-cost objective is not enough on its own. Two strategies can have the same $\mathbb{E}[J]$ and very different distributions: one with outcomes tightly grouped around the mean, the other usually cheap and occasionally catastrophic. The expectation averages the catastrophe into a small number and the optimizer becomes indifferent between the two. A risk-aware objective adds a measure of the bad outcomes,

$$\min_{u} \; \mathbb{E}[J] + \lambda \, \mathrm{Risk}(J), \tag{20.25}$$

with $\lambda \geq 0$ the price of risk. A common choice of Risk is the conditional value at risk. For a level q between 0 and 1, CVaR_q is the average loss over the worst $(1-q)$ fraction of cases, and the value at risk VaR_q is the threshold that begins that fraction, the q-quantile of the loss (the level is written q here because α denotes the recirculation fraction). At $q = 0.95$ the measure focuses on the worst 5 % of outcomes.

Consider two operating plans for a day. Plan A runs with generous margins: its energy cost is \$9800, \$10,000 or \$10,200 with probabilities 0.25, 0.5 and 0.25, an expected cost of \$10,000, with virtually no thermal risk. Plan B runs close to its limits: its energy cost is \$9000, and with probability 0.01 a thermal excursion forces a service interruption that adds \$50,000. Its expected cost is $0.99 \times 9000 + 0.01 \times 59{,}000 = \$9500$. An expected-cost optimizer picks B, which is \$500 cheaper. The worst 5 % of B's outcomes consist of the 1 % at \$59,000 and 4 % at \$9000, so

$$\mathrm{CVaR}_{0.95}(B) = \frac{0.01 \times 59{,}000 + 0.04 \times 9000}{0.05} = \$19{,}000, \qquad \mathrm{CVaR}_{0.95}(A) = \$10{,}200, \tag{20.26}$$

the second because the worst 5 % of A lie wholly in its top outcome. The value at risk misses the difference: $\mathrm{VaR}_{0.95}(B) = \$9000$ is below $\mathrm{VaR}_{0.95}(A) = \$10{,}200$, because B's catastrophe lies beyond the 95 % quantile. With $\lambda = 0.1$ in equation (20.25), A scores $10{,}000 + 0.1 \times 10{,}200 = \$11{,}020$ and B scores $9500 + 0.1 \times 19{,}000 = \$11{,}400$; the risk-aware optimizer prefers A. The two tie at $\lambda = 500/8800 = 0.057$, so any operator who values a dollar of tail loss at more than about six cents of expected cost prefers A.

Figure 20.6 draws the upper tail of the two cost distributions. The horizontal axis is the cost of the day in thousands of dollars and the vertical axis the cumulative probability, shown only from 0.90 to 1; both curves enter from below the axis. The horizontal line marks $q = 0.95$, and the shaded band above it is the worst 5 % over which CVaR averages. A's curve rises to 1 in one step at \$10,200, which is also its CVaR. B's curve rises to 0.99 at \$9000, runs flat to \$59,000 and only then reaches 1: its last percent of probability lies far to the right. The red

dashed mark is B's CVaR. The figure shows why a quantile alone misses the tail and why its average does not: B crosses the 0.95 line to the left of A, but the average over the band lies far to the right.

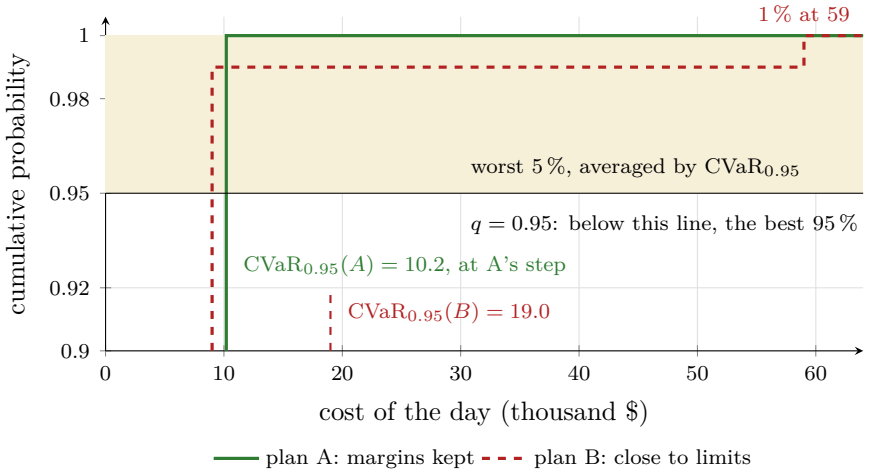


Figure 20.6: Upper tail of the cumulative distributions of two daily plans. Plan A: \$9.8, 10.0 and 10.2 thousand with probabilities 0.25, 0.5, 0.25. Plan B: \$9.0 thousand with probability 0.99 and \$59.0 thousand with probability 0.01. Their means are close (\$10.0 and \$9.5 thousand); their $\mathrm{CVaR}_{0.95}$ are 10.2 and 19.0 thousand. Illustrative values.

Principle 20.3. *Expected cost treats a rare catastrophe and a frequent nuisance of the same average as equivalent. A risk-aware objective prices the tail explicitly, through a chance constraint or a term such as* CVaR*, and can rightly prefer the more expensive plan.*

20.5 Dynamic, discrete and contingency structure

Three features make a data center optimization harder than a single static problem: decisions over time that must be revised as forecasts change, decisions that are on or off, and constraints that must hold in the states after failures as well as in the normal one. This section treats each and ends with contingency probabilities that depend on the condition of the equipment.

20.5.1 Model predictive control

Model predictive control (MPC), introduced in Chapter 10, is the form in which a dynamic optimization is operated. At each step it estimates the state, forecasts the IT load and the weather, solves an optimization over a horizon of H steps, applies the first action and updates; the chain is state estimation, forecasting, optimization and control, repeated. The example problem of this chapter is

$$\begin{aligned} \min_{u_{t:t+H}} \ & \sum_{k=t}^{t+H} \Big(p_k \, P_{\text{facility},k} \, \Delta t + \lambda_T \, \psi(T_k) \Big) \\ \text{subject to } & x_{k+1} = f(x_k, u_k, d_k), \quad T_{i,k} \le 35^{\circ}\text{C}, \quad P_{\text{facility},k} \le 100\,\text{MW}, \\ & \text{SOC}_k \ge 0.8, \quad \text{and the workload-completion constraints,} \end{aligned} \tag{20.27}$$

where $u_{t:t+H}$ is the sequence of decisions over the horizon, $T_{i,k}$ the inlet temperature of rack i at step k, and $\psi(T_k)$ a thermal penalty that grows as temperatures approach the limit, for

example $\psi(T_k) = \sum_i \max(0,\ T_{i,k} - 32^\circ\mathrm{C})^2$, priced at λ_T dollars per square kelvin. The penalty is the C_{thermal} of equation (20.12): it lets the optimizer use the band between 32 and 35°C at a price instead of treating it as either free or forbidden.

Every element of the chapter so far appears in equation (20.27). The price-weighted energy is equation (20.13), the thermal and electrical limits are those of Section 20.2, the reserve is equation (20.18), and the workload-completion constraints are the energy and deadline constraints of the batch job in equation (20.14). The state recursion couples the steps: a battery discharged now is unavailable later, and a job deferred now must run later. Because only the first action is applied and the problem is solved again with new measurements, the plan is revised as the forecasts of load and weather change, which is how a controller copes with the uncertainty of Section 20.4 even when each solve is deterministic.

20.5.2 Discrete decisions and mixed-integer optimization

Some decisions take continuous values. The fan speed can be 0.73 of its rating, as at the optimum of Figure 20.2, and a small change in it makes a small change in cost and temperature. Others are discrete: chiller 3 is on or off, a generator is running or not, a job runs in this hall or in that one. A problem with both kinds is a mixed-integer optimization.

Discrete decisions make the problem much more expensive to solve. With n on/off units there are 2^n combinations; each combination defines a different continuous problem; and the feasible set is no longer convex, so a solution that cannot be improved by small changes is not necessarily the best one. Mixed-integer solvers handle problems with hundreds of binary variables by branching on the discrete choices and bounding the continuous relaxations, but the solve time grows quickly with the number of units and time steps, which limits the horizon and the detail an MPC can afford when it re-solves every few minutes.

20.5.3 Chiller staging as unit commitment

The choice of which chillers to run is the standard discrete decision of a cooling plant. With an on/off variable $z_i \in \{0, 1\}$ for each of four chillers of capacity $\dot{Q}_i^{\max}$, the capacity constraint is

$$\sum_{i=1}^{4} \dot{Q}_i^{\max} z_i \ \geq\ \dot{Q}_{\text{req}}, \tag{20.28}$$

and the objective adds, to the electrical power of the running chillers, a startup cost $C_{\text{start},i}$ for each unit switched on. The problem also carries a minimum stable loading (a running chiller must carry at least a fraction of its capacity), a minimum runtime (a chiller once started must run for a minimum time, and once stopped must stay off for a minimum time) and an efficiency curve (the power of each unit as a function of its load). The structure is that of unit commitment in a power system, with chillers in the role of generating units and the cooling requirement in the role of the electrical demand.

Table 20.5 works an illustrative case. Chillers 1 and 2 have a capacity of 2.0 MW and an electrical power of $0.10 + 0.20\,\dot{Q}_i$ in megawatts at a load of $\dot{Q}_i$ megawatts; chillers 3 and 4 are newer, with 2.5 MW and $0.08 + 0.14\,\dot{Q}_i$. Every chiller must carry at least 0.6 MW when running. The requirement is $\dot{Q}_{\text{req}} = 5.2$ MW. Each row is a combination of running chillers; the columns give its capacity, the cheapest loading of its units, its electrical power and whether it survives the loss of its largest running unit. Within a combination the efficient units are loaded first, subject to the minimum loading of the others. The pair $\{3, 4\}$ is infeasible because 5.0 MW is less than 5.2 MW. The cheapest feasible combination is $\{1, 3, 4\}$ (or its twin $\{2, 3, 4\}$), at 1.024 MW: chiller 1 at its minimum and the two efficient units at 2.3 MW each, a plant coefficient of performance of $5.2/1.024 = 5.1$.

Table 20.5: Chiller staging for $\dot{Q}_{\text{req}} = 5.2\,\text{MW}$ with four chillers: capacity, cheapest loading, electrical power and survival of the loss of the largest running unit ($N{+}1$). Illustrative values.

Running	Capacity	Loading (MW)	Power	$N{+}1$ capacity
$\{3,4\}$	5.0 MW	infeasible	–	–
$\{1,3,4\}$	7.0 MW	$\dot{Q}_1 = 0.6$, $\dot{Q}_3 = \dot{Q}_4 = 2.3$	1.024 MW	4.5 MW: fails
$\{1,2,3\}$	6.5 MW	$\dot{Q}_1 = \dot{Q}_2 = 1.35$, $\dot{Q}_3 = 2.5$	1.170 MW	4.0 MW: fails
$\{1,2,3,4\}$	9.0 MW	$\dot{Q}_1 = \dot{Q}_2 = 0.6$, $\dot{Q}_3 = \dot{Q}_4 = 2.0$	1.160 MW	6.5 MW: holds

The last column adds the contingency constraint of the next subsection. If the plant must carry 5.2 MW after the loss of any running chiller, only the four-chiller combination qualifies, at 1.160 MW. The redundancy costs 0.136 MW, 13 % of the cheapest plant power, about \$196 a day at \$60 per MWh. Over a day the problem repeats at every step with a different requirement, and the startup cost and the minimum runtimes couple the steps: it can be cheaper to keep an inefficient chiller running through a short dip in load than to stop it and start it again an hour later.

20.5.4 Contingency-constrained optimization

The post-contingency constraint of equation (20.9) generalizes to every part of the facility. For a list of contingencies $c = 1, \dots, N_c$ (the loss of a chiller, a UPS module, a utility feed, a pump, an air handler), the optimization requires

$$g\big(x^{(c)}, u^{(c)}\big) \le 0 \quad \text{for } c = 1, \dots, N_c, \tag{20.29}$$

where $x^{(c)}$ is the state the facility would reach after contingency c and $u^{(c)}$ the corrective actions allowed in it, such as starting a standby chiller or capping IT power. The optimization therefore runs across many possible future states at once: the normal one and one per contingency, each with its own copy of the facility model and its own constraints.

The size of the problem grows with the number of contingencies. With N_c contingencies, a facility model of n state variables becomes $(N_c + 1)\,n$ variables, and with an MPC horizon of H steps the count is multiplied again. The list is usually limited to single failures, the $N{-}1$ criterion of power-system security, and the corrective actions are restricted to those that can be taken within the time the thermal mass of the facility allows (Chapter 8).

20.5.5 Risk-weighted and condition-dependent contingencies

The $N{-}1$ criterion treats every single contingency alike: each must be survived, whatever its probability. Contingencies are not alike. Consider three, with probabilities per day and consequences if the facility is unprotected: c_1, the trip of an air handler, with $\Pr(c_1) = 0.1$ and a loss $L_1 = \$20{,}000$ of throttled compute; c_2, a rupture in a secondary chilled-water loop, with $\Pr(c_2) = 10^{-4}$ and $L_2 = \$500{,}000$; and c_3, the trip of chiller 3, with $\Pr(c_3) = 0.01$ and $L_3 = \$50{,}000$. Protecting against each, by running a standby air handler, by running a redundant loop, or by staging chiller 4 early, costs an illustrative \$150, \$1600 and \$800 a day. The risk of each contingency is the probability times the loss,

$$\mathcal{R}_i = \Pr(c_i)\,L_i, \qquad \mathcal{R}_1 = \$2000, \quad \mathcal{R}_2 = \$50, \quad \mathcal{R}_3 = \$500, \tag{20.30}$$

and a risk-weighted rule protects against a contingency when its risk exceeds the cost of protection: c_1 is protected and c_2 and c_3 are accepted for the day. The deterministic $N{-}1$

rule would pay \$2550 a day to protect all three; the risk-weighted rule pays \$150. Treating c_1 and c_2 alike is unnecessarily conservative, because their probabilities differ by a factor of a thousand. The rule here is risk-neutral; for a contingency whose consequence is a critical-service interruption, the consequence-based tolerance of Section 20.4.4 applies instead, and the protection is bought whenever the probability exceeds ϵ_S.

The probabilities need not be fixed. If chiller 3 shows degradation, a rising condenser approach temperature or a vibration trend, the probability of its trip is a probability given the measurements, $\Pr(c_3 \mid y)$, estimated from the condition of the equipment as in Chapter 16. With an illustrative posterior of 0.6 on the hypothesis H_1 that the chiller is degraded, a trip probability of 0.15 per day if it is degraded and the base rate 0.01 if it is not,

$$\Pr(c_3 \mid y) = 0.6 \times 0.15 + 0.4 \times 0.01 = 0.094, \tag{20.31}$$

and the risk becomes $0.094 \times 50{,}000 = \$4700$, above the \$800 cost of protection. The optimizer now stages chiller 4, shifts load away from chiller 3 and preserves margin. The same facility with the same redundancy makes a different decision because the condition of one unit changed, which a fixed redundancy rule cannot do.

Figure 20.7 places the three contingencies on a map. The horizontal axis is the probability per day and the vertical axis the consequence in dollars, both logarithmic. The diagonal lines are lines of constant risk $\Pr \times L$ at \$100, \$1000 and \$10,000. Each dot is a contingency (the open dot is c_3 at its base rate), and the arrow moves c_3 from its base probability to its condition-dependent one. The figure shows the decision as a position: a contingency is worth protecting when the risk line through it lies above its protection cost. c_1 sits near the \$2000 line against a cost of \$150; c_2 sits on the \$50 line against \$1600; and c_3 moves from \$500 to \$4700 against \$800, crossing its protection cost when the diagnosis changes, without any change in its consequence.

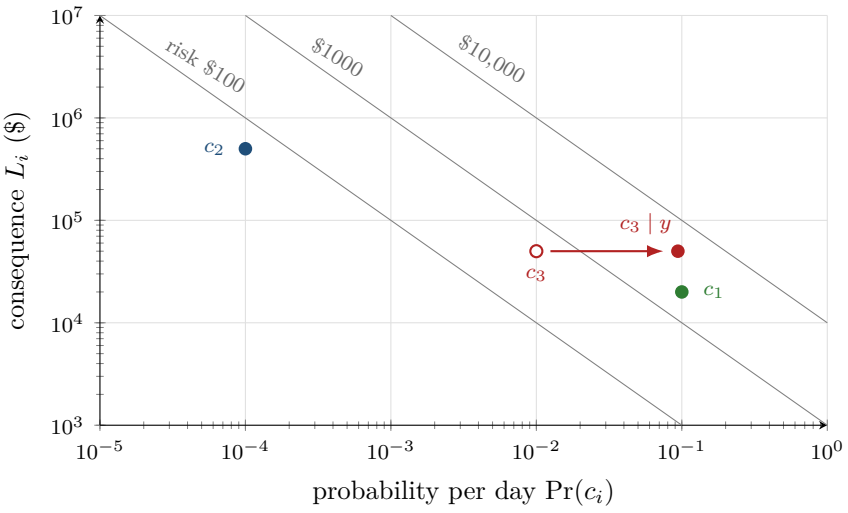


Figure 20.7: Contingencies on a probability–consequence map. Lines of constant risk $\Pr(c_i)L_i$. c_1, air handler trip (0.1, \$20,000, risk \$2000); c_2, loop rupture (10^{-4}, \$500,000, risk \$50); c_3, chiller trip (0.01, \$50,000, risk \$500), moved by the update of equation (20.31) to 0.094 and risk \$4700. Illustrative values.

Principle 20.4. *A contingency should be weighted by its probability and its consequence, not counted. Risk weighting avoids paying for protection against improbable minor events, and condition-dependent probabilities let the optimizer buy protection when a unit's measured condition makes its failure likely.*

20.6 The place of optimization in operation

An optimizer is only as good as the estimates it acts on and only as useful as the decisions it can get carried out. This section places it in the chain from data to action, describes how its recommendations reach an operator, extends it to investment decisions, collects the chapter into one formulation, and states its relation to the facility's existing controls.

20.6.1 Inference constrains optimization

An optimizer exploits whatever its model says. If the model says that the airflow through a rack is 1.5 kg/s and the optimizer treats that number as exact, it raises the supply air to 28.3°C, the value at the estimate in Figure 20.5, and the rack is safe only if the estimate is right. When the airflow is poorly identifiable (Chapter 15), when several airflows are equally consistent with the measurements, optimizing aggressively on the estimate $\hat{m}$ as if it were exact is dangerous: the optimizer moves the facility exactly in the direction in which the model is least supported.

There are three correct responses. The first is to carry the uncertainty into the optimization, as a chance constraint with the posterior's standard deviation, which backs off automatically where the estimate is weak. The second is to operate conservatively in the directions the data do not support, holding the setpoints that experience has shown to be safe. The third is to request a better measurement, which narrows the posterior and moves the safe point toward the limit, the subject of Chapter 19. In each case the inference layer constrains the optimization layer: what the optimizer may do depends on what the estimator can support.

20.6.2 The decision hierarchy

The order of the stages is therefore fixed. Measurements are checked for credibility and the quantities of interest for identifiability; the state is estimated with its uncertainty; the disturbances are forecast; the risk of each constraint is computed; the optimizer chooses an action; and the action reaches the plant through its controls. Figure 20.8 draws the chain. The top row runs left to right from the measurements to the forecast and the second row right to left from the risk to the setpoints; the bottom row is the physical plant with its low-level controls and safety interlocks, whose measurements return to the start of the chain. The operator sits between the optimizer and the setpoints. The figure shows each stage consuming the output of the one before it, which is the sense of the rule that optimization should not sit on untrustworthy inference: an error in an early box propagates to every box after it, and the optimizer, which seeks the boundary of the feasible set, amplifies it.

20.6.3 Recommendations and the operator

In most facilities the optimizer does not act directly. It recommends, and an operator approves or rejects. A recommendation states the action, its benefit and its risk, for example: *raise the chilled-water setpoint from* 7°C *to* 9°C*; predicted cooling power* $-8\,\%$*; predicted* $\Pr(\max_i T_{\mathrm{in},i} > 35^\circ\mathrm{C}) = 0.003$. The operator can check the probability against the facility's tolerance, here $0.003 < \epsilon_T = 0.01$, and against what is known that the model does not know, such as maintenance in progress. This human-in-the-loop arrangement is the realistic first deployment of an optimizer in a facility that has run on fixed setpoints: it builds trust in the recommendations before any are executed automatically, and it keeps a person accountable for each change.

A fuller output covers a coming period. Table 20.6 gives an example for 16:00 tomorrow, when the forecast shows a high IT load and a high wet-bulb temperature, and chiller 3 is scheduled out of service. The first row is the baseline: with no action, the probability of a thermal violation is 0.14. The second row is the recommended set of actions and their combined result: shift 0.6 MW of flexible workload away from hall B, start chiller 4 before 15:30 so that

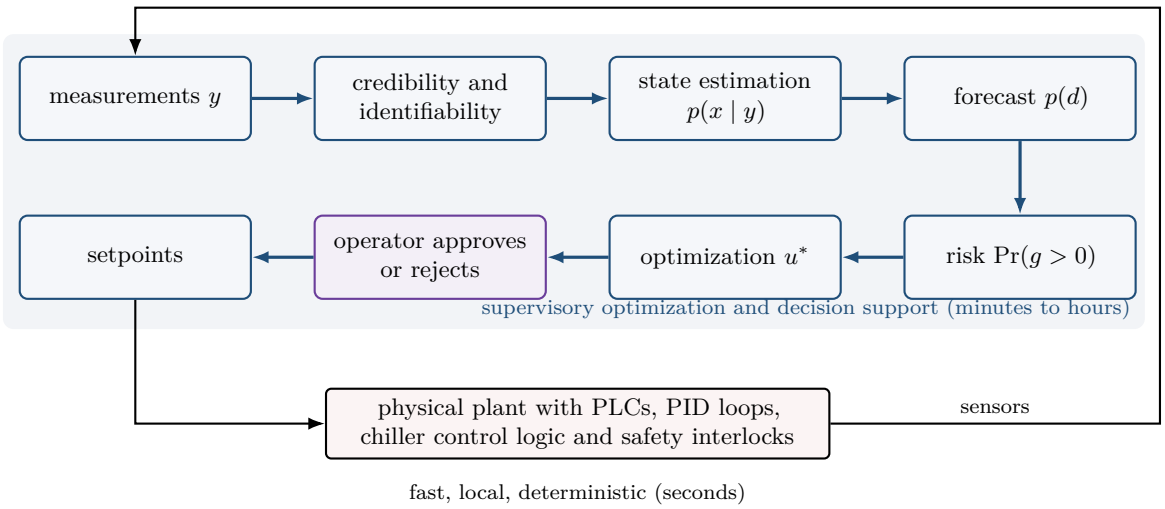


Figure 20.8: The decision hierarchy. Inference (top row) precedes risk and optimization (second row); the operator approves the recommended action before it reaches the setpoints of the low-level controls, which remain in charge of the plant and its safety interlocks.

it is at full output when the load peaks, and keep the chilled-water setpoint at 7°C rather than raising it as on a normal day. Together they bring the probability to 0.008, within the 0.01 tolerance, at a cost of 0.35 MW of additional facility power. Every number in the table is a calculated or inferred quantity, a prediction from the state estimate, the forecasts and the model; none is measured, and the report says so.

Table 20.6: An example of recommended actions for 16:00 tomorrow: high IT load, high wet-bulb temperature, chiller 3 unavailable. Illustrative values.

Situation or action	Pr(thermal violation)	Additional facility power
baseline, no action	0.14	–
recommended, applied together: (1) shift 0.6 MW of flexible workload from hall B; (2) start chiller 4 before 15:30; (3) keep the chilled-water setpoint at 7°C	0.008	+0.35 MW

20.6.4 Investment questions

The same machinery answers planning questions. A new chiller, an additional air handler, an airflow modification such as containment, a larger UPS or additional sensors each change the facility model, its constraints or the uncertainty of its estimates. For each candidate, the optimizer is run over a representative year of future operation and the result is compared with the present facility: the energy saved, the capacity gained at the same risk, and the risk removed at the same capacity, set against the capital cost. The comparison is between optimized operations, not between nameplates, because an investment's value is what it lets the optimizer do.

Additional sensors are an investment of this kind, and Figure 20.5 prices one. Suppose a set of airflow measurements narrows the standard deviation of the rack airflow estimate from 0.105 to 0.05 kg/s. At $\epsilon_T = 0.01$ the highest supply temperature rises from 23.1 to 26.1°C, 3.0 K without any change to the cooling equipment. At the 4 % per kelvin of equation (20.4)

and a 1.6 MW chiller plant, that is about 0.19 MW, and over a year at \$60 per MWh about $0.19 \times 8760 \times 60 = \$100{,}000$, illustrative but of the order of the cost of the sensors. Measurement and equipment compete for the same capital on the same basis.

20.6.5 The compact formulation and control as repeated optimization

The chapter collects into one problem,

$$\begin{aligned} &\min_{u} \ \mathbb{E}\big[\, C_{\text{energy}} + C_{\text{compute}} + C_{\text{wear}} + C_{\text{failure}} \,\big] \\ &\text{subject to } x_{k+1} = f(x_k, u_k, d_k, \theta), \\ &\qquad \Pr(T_i > T_{\max}) \le \epsilon_T, \quad \Pr(P_{\text{facility}} > P_{\text{grid,max}}) \le \epsilon_E, \\ &\qquad \text{SOC} \ge \text{SOC}_{\text{reserve}}, \quad W_{\text{served}} \ge W_{\text{required}}, \end{aligned} \tag{20.32}$$

where C_{compute} is the cost of deferred, slowed or migrated work (the C_{perf} of equation (20.12)) and C_{failure} the expected cost of failures, including the contingency risks of equation (20.30). The expectation is over the forecast disturbances, the uncertain parameters and the posterior state. This is a stochastic optimal-control problem: a dynamic model, an expected cost, probabilistic constraints set by consequence, and hard constraints where no risk is accepted.

Control and optimization are then two views of the same thing. An optimization returns a decision u^* for the conditions it was given. Control adds the system dynamics, feedback, repeated updating and execution in real time: it recomputes and applies $u^*(t)$ as the conditions change. Model predictive control is exactly this repetition of equation (20.32) over a receding horizon, and the difference between an optimizer and a controller lies in how often the problem is solved and whether the answer is applied automatically.

20.6.6 The supervisory layer

Equation (20.32) is not a replacement for the controls a facility already has. Programmable logic controllers (PLCs), the industrial computers that run equipment sequences, the PID loops of Chapter 10, the chillers' own control logic and the safety interlocks that stop equipment on a dangerous condition are mature, fast and low-level. They act in seconds, are verified against known failure modes, and must keep the plant safe when everything above them is wrong. An optimizer that bypassed them would trade that reliability for an energy saving.

The optimizer belongs above them, as a supervisory optimization and decision-support layer that works in minutes to hours and acts through setpoints, staging plans and workload recommendations. Such a layer provides uncertain state estimates with their standard deviations, capacity stated as a risk rather than a nameplate, contingency analysis of the post-failure states, what-if evaluation of candidate actions, recommended operating actions with their predicted benefit and risk, and recommendations for the next measurement to take. Its outputs are the rows of Table 20.6, and its inputs are the inference of the top row of Figure 20.8.

> **Principle 20.5.** *Optimization belongs between inference and control: it must not act on estimates the data do not support, and it should act on the plant through the low-level controls and their interlocks, as a supervisory layer that recommends and explains its actions.*

20.7 What to retain

Definition 20.1 (Data center optimization)**.** Data center optimization is the choice of how to operate the IT, cooling and electrical resources together, $\min_u \mathbb{E}[J]$ over the dynamics $x_{k+1} = f(x_k, u_k, d_k, \theta)$, subject to thermal, electrical, cooling and computing constraints in the

normal and post-contingency states, with risk tolerances set by consequence; under uncertainty its question is which action gives the best performance while keeping the physical risk acceptably low.

Table 20.7 collects the chapter's objects with the subsection that defines each.

Table 20.7: The objects of data center optimization.

Object	Definition	Section
$\min_u J$ s.t. $g \le 0$, $x_{k+1} = f(\cdot)$	the generic problem	20.1.1
$u = [u_{\text{cool}}, u_{\text{IT}}, u_{\text{elec}}]$	the three groups of decisions	20.1.2
$\min P_{\text{cool}}$ s.t. $T_{\text{in},i} \le T_{\max}$	the cooling-energy problem; the optimum lies on an active constraint	20.1.3
$T_{\text{in},i} \le T_{\max} - \Delta T_{\text{margin}}$	thermal constraint with margin	20.2.1
$P^{(c)}_{\text{UPS},j} \le P_{\text{UPS,max},j}$	post-contingency electrical constraint	20.2.2
$W_{\text{served}} \ge W_{\text{required}}$, $t_{\text{lat}} \le t_{\text{lat,max}}$	computing constraints	20.2.3
$J = \sum_m w_m C_m$, $C_{\text{energy}} = \sum_t p_t P_{\text{facility},t} \Delta t$	the weighted objective and the cost of energy	20.2.4
$\int P_{\text{job}}\,\mathrm{d}t = E_{\text{job}}$	flexible-load scheduling against a price	20.3.2
$\text{SOC}_t \ge \text{SOC}_{\text{reserve}}$	reserve constraint; its dual value prices resilience	20.3.5
$g(x, u, \theta) \le 0\ \forall \theta \in \Theta$	robust optimization; conservative	20.4.2
$\Pr(g \le 0) \ge 1 - \epsilon$; back-off $z_{1-\epsilon}\sigma$	chance constraint	20.4.3
ϵ_T, ϵ_E, ϵ_S	risk tolerances set by consequence	20.4.4
$\mathbb{E}[J] + \lambda\,\text{CVaR}_q(J)$	expected cost plus tail risk	20.4.5
model predictive control	repeated horizon optimization, first action applied	20.5.1
$\sum_i \dot{Q}_i^{\max} z_i \ge \dot{Q}_{\text{req}}$, $z_i \in \{0,1\}$	chiller staging as unit commitment	20.5.3
$g(x^{(c)}, u^{(c)}) \le 0$, $c = 1, \ldots, N_c$	contingency-constrained optimization	20.5.4
$\mathcal{R}_i = \Pr(c_i \mid y) L_i$	risk-weighted, condition-dependent contingencies	20.5.5
decision hierarchy	inference before risk, risk before optimization	20.6.2
equation (20.32)	stochastic optimal control	20.6.5
supervisory layer	decision support above PLCs, PID loops and interlocks	20.6.6

20.8 Exercises

Exercise 20.1. The batch job of Section 20.3.2 has its power limit lowered to $P_{\text{job}}^{\max} = 2\,\text{MW}$. (a) How many hours does it need, and which hours does the solution of equation (20.14) use with the prices and overhead factors of that subsection? (b) Compute its cost and compare it with running on arrival from 16:00. (c) The deadline is moved to 05:00. Which hours remain available, what is the new optimal cost, and what is the value of the three hours of deadline that were lost?

Exercise 20.2. (a) A facility power forecast has $\sigma = 2\,\text{MW}$ and the limit is 100 MW at $\epsilon_E = 0.001$. Compute the highest planned mean draw. (b) The rack of equation (20.22) has a delivered airflow estimate of 1.5 kg/s with $\sigma = 0.08\,\text{kg/s}$. Compute the first percentile of the

airflow and the highest supply temperature that satisfies $\Pr(T_{\text{in}} > 35^{\circ}\text{C}) \le 0.01$. (c) Compute the robust supply temperature if the plausible set extends to 1.0 kg/s, and the probability of that airflow under the normal estimate. (d) Using 4 % of a 1.6 MW chiller plant per kelvin, estimate the power the robust controller spends beyond the chance-constrained one.

Exercise 20.3. A third plan C has a daily cost of \$9200 with probability 0.95 and \$25,000 with probability 0.05. (a) Compute its expected cost, $\text{VaR}_{0.95}$ and $\text{CVaR}_{0.95}$. (b) With $\lambda = 0.1$ in equation (20.25), rank plans A, B and C of Section 20.4.5. (c) Find the value of λ at which C and A tie, and say which plan an expected-cost optimizer would choose.

Exercise 20.4. For the chillers of Table 20.5, take $\dot{Q}_{\text{req}} = 4.4\,\text{MW}$. (a) List the feasible combinations and find the cheapest, with its loading and electrical power. (b) Repeat with the $N{+}1$ requirement that 4.4 MW is still covered after the loss of the largest running unit, and state the cost of the redundancy in megawatts and percent. (c) A diagnosis places a posterior of 0.4 on degradation of chiller 3, with a trip probability of 0.2 per day if degraded and 0.01 if not. Compute $\Pr(c_3 \mid y)$ and its risk at $L_3 = \$50{,}000$, and say whether protection at \$800 a day is worth buying.

20.9 Connection to the TERRA framework

TERRA supplies the model, the posterior and the risk that every formulation of this chapter consumes, and the supervisory layer of Section 20.6.6 is the role it plays above a facility's own controls.

- **The constraint set is the graph.** The facility model f of equation (20.1) is three coupled conservation networks (electrical power on buses, air and water mass on plenums, aisles and loops, thermal energy on racks and coils). The thermal, electrical and cooling constraints of Table 20.2 are bounds on node states, edge flows and component ratings, so a decision the graph cannot carry is infeasible by construction.
- **Decisions are component states and setpoints.** Fan and pump speeds, valve openings, the chilled-water setpoint, the on/off states z_i of chillers, the battery power with its SOC recursion and the IT power at each rack node are the inputs of components with efficiencies, capacities and bounds; forward simulation under a candidate u is the prediction step of the MPC of equation (20.27).
- **Chance constraints are propagated posteriors.** The airflow standard deviation that sets the back-off in Figure 20.5 is the posterior of an inferred edge flow; propagating samples of the posterior through the graph under each action gives $\Pr(T_{\text{in}} > T_{\max})$ and the probability that required cooling exceeds available cooling in Table 20.3.
- **Contingencies are graph modifications weighted by the posterior.** Removing a chiller or a UPS component and re-solving the balances gives the post-contingency state $x^{(c)}$, and the ranked posterior over hypotheses for a degrading unit gives the condition-dependent $\Pr(c_3 \mid y)$ of equation (20.31).
- **Identifiability gates the optimizer, and measurements are investments.** An inferred quantity whose posterior is wide widens the back-off rather than being optimized on as if exact, and the next measurement is chosen for the information it adds, valued, as in Section 20.6.4, by the supply temperature it releases.

In summary, a data center optimization is a stochastic optimal-control problem whose constraints are the facility's conservation laws and ratings, whose uncertainty is the posterior

over unmeasured states and parameters, and whose risk is the probability that a constraint is violated in the normal or a post-contingency state; a graph model with posterior inference supplies all three, and the optimizer, kept above the facility's own controls, turns them into recommended actions with their risks. Chapter 21 extends the decision beyond the fence of the facility, to its interaction with the electrical grid and the use of its load as a flexible resource.

Chapter 21

Grid Interaction and Flexible Load

21.1 Flexibility of a data center facility

The preceding chapters treated the electricity supply as a given: the utility delivers whatever the facility draws, and the facility's task is to convert it into computation and remove the resulting heat. A large data center is also one of the largest single loads on a distribution or transmission system, and it contains equipment, a battery, generators, deferrable computation and a cooling plant with thermal mass, that can change how much it draws at short notice. This chapter treats the data center as a potential participant in the operation of the power grid rather than only as a load. Its question is

how much electrical flexibility can a data center provide to the grid without compromising its computing service or its resilience?

The answer is not a number of megawatts. It is a trajectory in time with a magnitude, a duration, a speed of response, a recovery afterward and a confidence, set by the state of the battery, the thermal margin of the rooms, the deadlines of the computing work and the reserve the facility must keep for its own outages. The chapter defines flexibility, goes through each resource that provides it, shows how a commitment to the grid is measured against a counterfactual and forecast as a probability, and ends with the reserve that resilience requires, the price and carbon signals that drive flexibility, and a compact formulation of the whole problem.

Jargon: flexibility, demand response, curtailment, rebound, baseline, firm offer, frequency response, droop, behind the meter, aggregator, virtual power plant, carbon intensity

Flexibility is the ability of a facility to change its net electrical demand on request, relative to what it would otherwise have drawn. *Demand response* is a program or market in which a grid operator, a utility or an aggregator asks or pays consumers to change their demand for a period, usually to reduce it at times of scarcity. *Curtailment* is a reduction of demand below its normal level. *Rebound*, also called payback, is the extra demand after a curtailment, when deferred work runs, batteries recharge or warm rooms are cooled back down. The *baseline* is the demand the facility would have drawn without the request; it is a counterfactual and is never measured. A *firm* offer is a quantity the facility can deliver with a stated high probability, as opposed to its expected or best-case value. *Frequency response* is an automatic change of power in proportion to the deviation of the grid frequency from its nominal value; the proportional law is called *droop*. *Behind the meter* describes equipment on the customer's side of the utility meter, whose effect the grid sees only through the net power at that one connection point. An *aggregator* combines many small distributed energy resources (thermostats, electric vehicles, batteries, water heaters) into one controllable block, often called a *virtual power plant*. The *carbon intensity* of grid electricity, $\mathrm{CI}(t)$, is the mass of carbon dioxide emitted per unit of electricity generated at time t, in tonnes per megawatt hour.

Key idea

The flexibility of a data center is a trajectory of changed demand with a magnitude, a duration, a ramp, a rebound and a confidence. What can be committed to the grid is the part that remains safe with high probability under the forecast conditions, after the backup reserve, the thermal margins, the electrical limits and the computing deadlines have been respected.

This section writes the net grid power of a facility, defines flexibility in each direction, lists the attributes a flexibility offer must state, and shows that an offer is a trajectory with a rebound.

21.1.1 The facility as a controllable load

The power that crosses the utility meter is the sum of what the facility consumes minus what it supplies to itself. Writing the consumption as IT power, cooling power and auxiliary power (lighting, offices, losses not counted elsewhere), and the on-site supply as the discharge power of the battery and the output of the generators, the net grid power is

$$P_{\mathrm{grid}}(t) = P_{\mathrm{IT}}(t) + P_{\mathrm{cool}}(t) + P_{\mathrm{aux}}(t) - P_{\mathrm{batt}}(t) - P_{\mathrm{gen}}(t), \tag{21.1}$$

with $P_{\mathrm{batt}} > 0$ on discharge and $P_{\mathrm{batt}} < 0$ while charging. The sum of the first three terms is the facility power P_{facility} of Chapter 11; the grid sees only the difference $P_{\mathrm{facility}} - P_{\mathrm{batt}} - P_{\mathrm{gen}}$. Sign conventions differ between sources (some write the battery as a load with $P_{\mathrm{batt}} > 0$ on charge); the one used here makes discharge reduce grid demand, which is the direction that matters for most grid requests.

Each term of equation (21.1) can be changed by a decision. IT power changes when computing work is deferred, moved to another site, or slowed by a power cap. Cooling power changes when setpoints are raised or chillers are staged, and the thermal mass of the rooms and of the chilled water carries the facility through the period in which less heat is removed. The battery can discharge or charge. The generators can take part of the load. Each lever has

limits, and together they define a feasible range of grid power at every moment,

$$\underline{P}(t) \le P_{\text{grid}}(t) \le \overline{P}(t), \tag{21.2}$$

where $\underline{P}(t)$ is the lowest net demand the facility can reach at time t without violating any of its own constraints and $\overline{P}(t)$ the highest. Both bounds move with the state of the facility: with the state of charge of the battery, the temperatures in the rooms, the work in the queue and the weather.

Figure 21.1 draws equation (21.1) as a network. The box on the left is the utility grid and the circle is the meter, the single point at which the grid observes the facility. The thick vertical line is the facility's internal electrical bus. The five boxes on the right are the terms of the equation: three loads that draw from the bus, the battery that can draw or supply, and the generators that supply. The notes beside each box name the flexibility it offers. Everything inside the dashed outline is behind the meter: the grid sees none of it individually, only the net power through the circle.

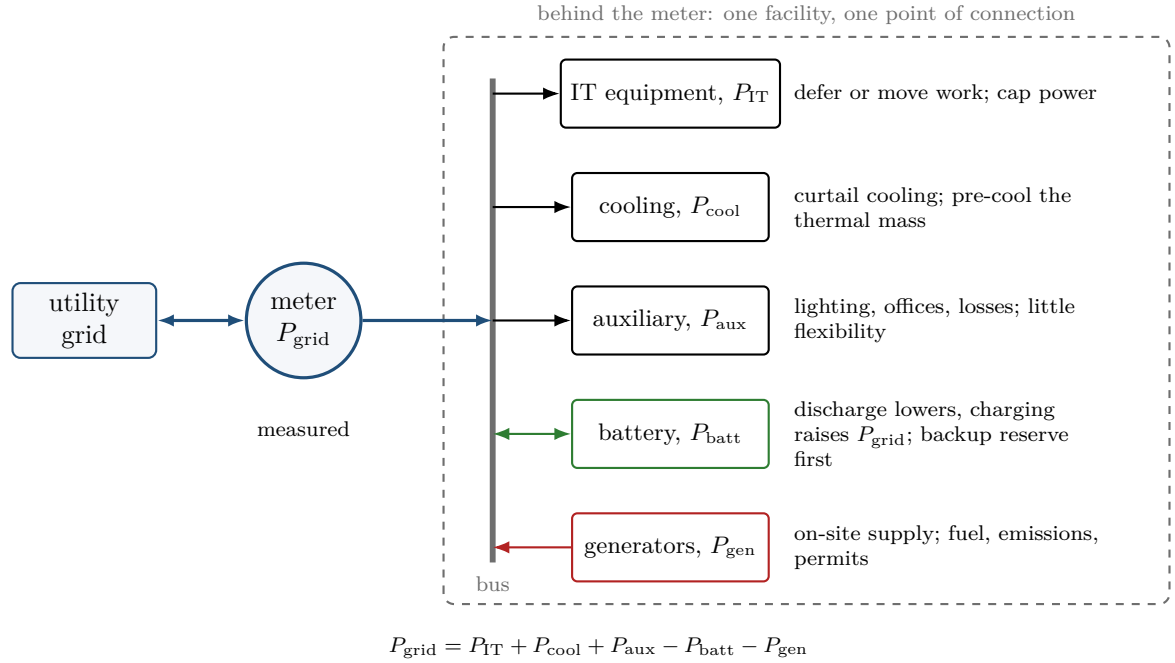


Figure 21.1: The data center as a portfolio of resources behind one meter, equation (21.1). Three loads draw from the facility bus, the battery draws or supplies, the generators supply; the grid observes only the net power P_{grid} at the meter. The notes name the flexibility each resource offers.

21.1.2 Directional flexibility

Flexibility is measured from the baseline, the demand the facility would draw without any request. Write the baseline as $P^{\text{base}}(t)$. The downward and upward flexibilities are

$$F^{\downarrow}(t) = P^{\text{base}}(t) - \underline{P}(t), \qquad F^{\uparrow}(t) = \overline{P}(t) - P^{\text{base}}(t), \tag{21.3}$$

the distances from the baseline to the two ends of the feasible range of equation (21.2). For an illustrative facility with a baseline of 50 MW that can reduce its demand to 45 MW, the downward flexibility is $F^{\downarrow} = 5\,\text{MW}$. If the same facility can raise its demand to 53 MW, by charging its battery or by advancing flexible jobs that would otherwise run later, its upward flexibility is $F^{\uparrow} = 3\,\text{MW}$.

The two numbers are different because different resources provide them. Downward flexibility comes from discharging the battery, deferring or capping computation, curtailing cooling and starting generators. Upward flexibility comes from charging the battery, which requires that it not be full, from running work early, which requires that work be queued, and to a small extent from cooling below the normal setpoint. A battery at a state of charge of 0.7 can absorb energy; one at 1.0 cannot. Grid operators value the two directions differently as well: downward flexibility helps at times of scarcity, upward flexibility at times of surplus, such as midday solar generation or windy nights. Flexibility is directional, and an offer must say which direction it is in.

21.1.3 The attributes of a flexibility offer

A statement such as "the facility can provide 5 MW" is incomplete. It leaves open five questions, and each answer changes what the statement is worth.

- *Magnitude*: how many megawatts, in which direction.
- *Duration*: for how long the change can be held. A battery with 1 MWh above its reserve can hold 5 MW for 12 minutes, not for an hour.
- *Ramp rate*: how quickly the change is reached. Some grid services need the full response within a second, others within minutes.
- *Rebound*: what extra demand follows the event, and for how long, when deferred work runs and stored energy is replaced.
- *Availability*: under which conditions and with what confidence the change can be delivered. A thermal curtailment that is safe on a cool morning may be impossible on a hot afternoon.

The confidence is the attribute most often left out. An offer of 5 MW that the facility can deliver nine times in ten is a different product from one it can deliver ninety-nine times in a hundred, and the difference is what Section 21.5 is about.

21.1.4 Flexibility as a trajectory

An illustrative flexibility product combines the five attributes: a curtailment of 5 MW, reached in under 30 s, held for 30 minutes; afterward, up to 2 MW of additional demand for 20 minutes to complete deferred jobs or recharge batteries. The product is not a scalar. It is a trajectory of the change in grid power,

$$\Delta P(t) = P_{\text{grid}}(t) - P^{\text{base}}(t), \tag{21.4}$$

negative during the curtailment and positive during the rebound.

Figure 21.2 plots this product. The horizontal axis is time in minutes from the start of the event; the vertical axis is the net grid power. The dashed line is the baseline of 50 MW. The solid line is the grid power with the product delivered: it falls to 45 MW within the first half minute, stays there for 30 minutes, rises to 52 MW for 20 minutes, and returns to the baseline. The blue area between the solid line and the baseline is the energy not drawn during the event, $5\,\text{MW} \times 0.5\,\text{h} = 2.5\,\text{MWh}$; the red area is the rebound energy, $2\,\text{MW} \times (20/60)\,\text{h} = 0.67\,\text{MWh}$. The figure shows that an offer specifies the whole shape: the depth, the width, the steepness of the first edge and the size of the hump that follows. Section 21.5.5 returns to the energy balance between the two areas.

Principle 21.1. *Flexibility is a trajectory $\Delta P(t)$, not a number: magnitude and direction, duration, ramp rate, rebound and the confidence of delivery all belong to the offer.*

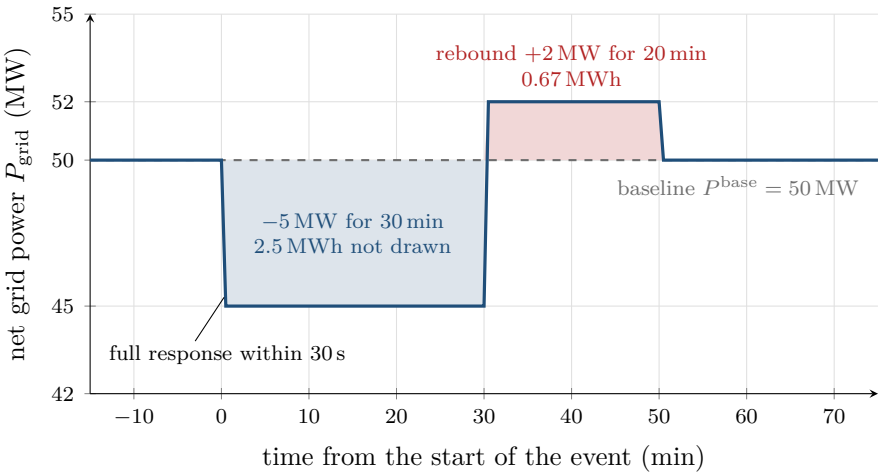


Figure 21.2: An illustrative flexibility product as a trajectory, $\Delta P(t) = P_{\text{grid}}(t) - P^{\text{base}}(t)$ (equation (21.4)): −5 MW reached within 30 s and held for 30 minutes, then +2 MW of rebound for 20 minutes. Blue: energy not drawn; red: rebound energy.

21.2 Electrical and IT resources

This section goes through the resources that act on the electrical side of equation (21.1) directly: the battery, which is constrained by its backup duty and its state, and the computing load, which can be deferred against a deadline or capped at a cost in throughput. It ends with the cost of providing flexibility.

21.2.1 Battery discharge and net grid power

During a discharge the battery supplies part of the facility's consumption, and by equation (21.1) the grid supplies less,

$$P_{\text{grid}} = P_{\text{facility}} - P_{\text{batt}}, \tag{21.5}$$

with the generators off. A discharge of $P_{\text{batt}} = 5$ MW lowers grid demand by about 5 MW without any change in the computing load, the cooling or the temperatures; the small difference is the conversion loss of the battery's power electronics. That independence from the computing service is what makes the battery the first resource considered for grid services: it is fast, it is precise, and the IT equipment does not see it. Its state of charge evolves as in Chapter 20, $\text{SOC}_{t+1} = \text{SOC}_t - P_{\text{batt},t}\Delta t / E_{\text{usable}}$, with E_{usable} the usable energy and conversion losses omitted.

The battery of a data center is, however, not installed for the grid. It is the energy store of the uninterruptible power supply (Chapter 4), sized to carry the critical load from the moment the utility fails until the generators have started and taken over. Every grid service it provides draws on the same energy that backup requires.

21.2.2 The backup reserve

Take an illustrative battery with $E_{\text{usable}} = 10$ MWh supporting a critical load of $P_{\text{crit}} = 40$ MW. At a state of charge SOC its backup runtime is

$$t_{\text{backup}} = \frac{\text{SOC}\, E_{\text{usable}}}{P_{\text{crit}}}, \tag{21.6}$$

which is 10/40 h = 15 minutes when full. Discharging it for a grid event from SOC = 1.0 to 0.7, which at 5 MW takes 3 MWh/5 MW = 36 minutes, leaves 10.5 minutes of runtime. If a utility

outage follows the event before the battery has recharged, the facility meets it with 30 % less time for the generators to start, synchronize and take the load, and with less time for a second attempt if the first start fails.

The simplest protection is a floor on the state of charge,

$$\mathrm{SOC}(t) \geq \mathrm{SOC}_{\mathrm{reserve}}, \tag{21.7}$$

for example $\mathrm{SOC}_{\mathrm{reserve}} = 0.6$, which keeps 9 minutes of runtime at 40 MW, and only the energy above the floor is offered to the grid. The floor is a proxy. What resilience requires is that the runtime be long enough with high probability,

$$\Pr\big(t_{\mathrm{backup}} < t_{\mathrm{required}}\big) < \epsilon_B, \tag{21.8}$$

where t_{required} is the time the generators need and ϵ_B the accepted probability of a failed ride-through. The probability is over the uncertain quantities: the usable energy, which depends on the battery's state of health and is inferred rather than measured (Chapter 4), and the time the generators take, which depends on whether they start at the first attempt. Section 21.6.2 computes a reserve from equation (21.8) and shows that it moves with the risk of an outage.

21.2.3 State dependence of battery flexibility

The flexibility a battery can offer for an event of duration τ is limited by its power and by its energy above the reserve,

$$F^{\downarrow}_{\mathrm{batt}}(\tau) = \min\!\left(P^{\mathrm{lim}}_{\mathrm{batt}},\ \frac{(\mathrm{SOC} - \mathrm{SOC}_{\mathrm{reserve}})\,E_{\mathrm{usable}}}{\tau}\right), \tag{21.9}$$

where $P^{\mathrm{lim}}_{\mathrm{batt}}$ is the discharge power the facility allows for grid service, here an illustrative 5 MW (the converters are rated for the critical load, but the facility limits how much of that rating it exposes to the grid). In words, the battery's flexibility is a function of its state of charge, its state of health through E_{usable}, its power limit, its usable energy and the backup requirement through $\mathrm{SOC}_{\mathrm{reserve}}$.

The state dependence is large. With the reserve at 0.6 and a 30-minute event, a battery at $\mathrm{SOC} = 0.95$ holds $(0.95 - 0.60) \times 10 = 3.5$ MWh above the reserve, enough for 7 MW over half an hour, so it offers the full 5 MW. At $\mathrm{SOC} = 0.75$ it holds 1.5 MWh and offers 3 MW. The state of health enters twice. If the battery has aged to $\mathrm{SOH} = 0.9$, its usable energy is 9 MWh, and holding the same 9 minutes at 40 MW, that is 6 MWh, now needs $\mathrm{SOC}_{\mathrm{reserve}} = 6/9 = 0.667$. At $\mathrm{SOC} = 0.95$ the energy above the reserve is $(0.95 - 0.667) \times 9 = 2.55$ MWh, still enough for 5 MW over half an hour; at $\mathrm{SOC} = 0.75$ it is 0.75 MWh, enough for 1.5 MW, half of what the new battery offered at the same state of charge. The battery's flexibility is therefore recomputed continuously from the estimated state of charge and state of health (Chapter 14), not read from its nameplate.

Upward flexibility. The same battery offers upward flexibility by charging, limited by the room left below full. At $\mathrm{SOC} = 0.7$ there are 3 MWh of room: 3 MW for an hour, or 2 MW for an hour and a half. Upward and downward flexibility trade against each other: a high state of charge maximizes the one and removes the other.

21.2.4 Workload deferral and the deadline constraint

Computing work divides into two classes for this purpose. *Time-sensitive* work must run when it arrives: web services, financial transactions, online inference that answers user requests. Delaying it degrades a service that someone is waiting for. *Deferrable* work has a deadline,

not a start time: training of machine-learning models, batch analytics, scientific simulations, backups. If 2 MW of computing can be deferred for one hour, that 2 MW is a demand-response resource of one hour's duration, and the facility's electrical demand during that hour drops by 2 MW plus the cooling power that would have removed its heat (Chapters 2 and 11).

A deferrable job j is described by the energy it needs and its deadline. If job j needs $E_j = 2\,\text{MWh}$ before $d_j = 20{:}00$, then whatever its schedule,

$$\int_{t_0}^{d_j} P_j(t)\,\mathrm{d}t = E_j, \qquad 0 \le P_j(t) \le P_j^{\lim}, \tag{21.10}$$

where $P_j(t)$ is the power the job draws and $P_j^{\lim}$ the most it can draw (the servers allocated to it at full speed). The equation is a conservation of work: curtailing the job now does not remove its energy, it moves it later. With $P_j^{\lim} = 2\,\text{MW}$, the job needs one hour at full power and must start by 19:00 at the latest. Its slack at time t,

$$t_j^{\text{slack}}(t) = (d_j - t) - \frac{E_j^{\text{rem}}(t)}{P_j^{\lim}}, \tag{21.11}$$

with E_j^{rem} the energy still to be delivered, is the longest it can still be deferred. At 16:00 with nothing yet done, the slack is $4 - 1 = 3$ hours. If the job had been planned at 1 MW from 16:00 to 18:00 and an event asks for curtailment from 17:00 to 18:00, the job can stop at 17:00 with 1 MWh done and finish at 2 MW from 18:00 to 18:30, or at 1 MW from 18:00 to 19:00. Either way the curtailment creates load later: the rebound. A job with zero slack offers no flexibility at all, and a queue of jobs whose deadlines all fall inside the event window offers none either.

21.2.5 Power capping

A power cap is a limit on the electrical power a server or accelerator may draw, enforced by its firmware within seconds (Chapter 2). Take an illustrative cluster of 100 accelerator (GPU) racks at 100 kW each, 10 MW in all. Capping each rack at 90 kW removes

$$\Delta P = 100 \times (100 - 90)\,\text{kW} = 1\,\text{MW} \tag{21.12}$$

quickly, and the heat removed with it lowers cooling power as the plant responds, by about 0.2 MW more at an illustrative ratio of cooling power to IT power of 0.2. The cost is lost computational throughput: the jobs on those racks run more slowly. Because the speed of a processor grows less than proportionally with its power near the top of its range, a 10 % power cap typically costs less than 10 % of throughput; an illustrative 4 % is used later. Capping can be held for a long time, since nothing is depleted, and it has little rebound if the slowed work is simply finished later at the same energy per unit of work. Its penalty is in the service, not in the energy.

21.2.6 The service cost of flexibility

Every resource that provides flexibility has a cost, and the costs are of different kinds,

$$C_{\text{flex}} = C_{\text{compute}} + C_{\text{battery}} + C_{\text{thermal}} + C_{\text{recovery}}, \tag{21.13}$$

where C_{compute} is the value of delayed or slowed computation, C_{battery} the degradation and conversion loss of the discharge, C_{thermal} the cost of operating closer to the temperature limits (a higher probability of a thermal excursion, faster aging of equipment), and C_{recovery} the cost of the rebound (energy bought back, possibly at a higher price, and any operational disruption). The facility's decision is then

$$\max_{a} \left[\,\text{grid revenue}(a) - C_{\text{flex}}(a)\,\right] \quad \text{subject to the reliability and service constraints,} \tag{21.14}$$

over the actions a that provide the flexibility: which resources, how much of each, when.

The costs are usually stated per megawatt hour of curtailment, which puts the resources in a merit order. With illustrative values, battery discharge costs about \$50 per MWh (degradation of the cells plus a round-trip loss of about 10 % that must be bought back), workload deferral about \$40 per MWh (the value of an hour's delay to batch work), and power capping about \$80 per MWh (the 4 % throughput loss on a 10 MW cluster whose computation is valued at \$2000 per hour is \$80 per hour for 1 MW of curtailment). If the grid pays an illustrative \$300 per MWh of curtailment, all three are worth dispatching. The constraint in equation (21.14) matters more than the costs: a resource whose cost is low but whose use would breach the backup reserve or a computing deadline is not available at any price.

21.3 Thermal and generation resources and the portfolio

The second group of resources acts through the cooling plant and the generators. This section quantifies thermal flexibility as a duration set by temperature headroom, shows how pre-cooling extends it and how conditions limit it, discusses generators, and then combines all the resources into a portfolio with different speeds of response.

21.3.1 Cooling curtailment through thermal inertia

Cooling power can be reduced temporarily without reducing the heat the IT equipment produces. Take an illustrative facility whose cooling plant draws 8 MW and can be reduced to 6 MW for a short period, by raising chilled-water and supply-air setpoints and unloading chillers. The 2 MW reduction is shed at once. The heat that is no longer removed accumulates in the thermal mass of the facility (the air, the racks, the building structure and above all the chilled water in the pipes and tanks), and temperatures rise. By the heat balance of Chapter 8,

$$C_{\mathrm{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\mathrm{IT}} - \dot{Q}_{\mathrm{avail}}, \tag{21.15}$$

the temperature rises at a rate equal to the unremoved heat divided by the effective thermal capacitance C_{th}. For a period the temperatures stay within their limits. The facility is using its thermal mass as storage: thermal storage without a battery.

The rate of rise is a measured or inferred quantity, not a datasheet value. If, for example, the reduction leaves 10 MW of heat unremoved and the rack inlet temperatures are observed to rise at 0.4 K per minute, equation (21.15) implies an effective capacitance of $10^7\ \mathrm{W}/(0.4/60\ \mathrm{K/s}) = 1.5 \times 10^9\ \mathrm{J/K}$, the equivalent of about 360 tonnes of water. That number lumps every mass that exchanges heat with the air on the time scale of the event; it is a parameter to be estimated from past events (Chapter 14), and it is the parameter on which the duration of thermal flexibility depends.

21.3.2 Duration of thermal flexibility

The duration follows from the headroom to the limit and the rate of rise. With the time-to-limit of Chapter 8,

$$t_{\mathrm{limit}} = \frac{T_{\max} - T}{\mathrm{d}T/\mathrm{d}t}, \tag{21.16}$$

a rack at 24°C with a maximum of 32°C has 8 K of headroom; at 0.4 K per minute under reduced cooling, it reaches the limit in $8/0.4 = 20$ minutes. The cooling flexibility of this facility is therefore 2 MW for 20 minutes, 0.67 MWh of curtailment, and not 2 MW without qualification. A magnitude without a duration is meaningless for thermal flexibility, because the magnitude can be held only until the headroom is used.

Two refinements apply in practice. The rise under constant reduced cooling is not exactly linear: the temperature approaches a new equilibrium along an exponential (Chapter 8), so the initial slope overestimates the later rate, and equation (21.16) with the initial slope is conservative. And no operator runs a curtailment to the limit itself: stopping at 31°C keeps 1 K for the recovery, during which the temperature continues to rise until restored cooling catches up, and shortens the duration to $7/0.4 = 17.5$ minutes. Where the rise rate scales with the size of the curtailment, as it does when the unremoved heat is proportional to the cooling power shed, the product of magnitude and duration is roughly fixed: 2 MW for 20 minutes or 1 MW for 40 minutes.

21.3.3 Pre-cooling

The headroom can be enlarged before an expected event. If the facility cools its thermal mass below the normal operating temperature ahead of time, the event starts from a lower temperature and the limit is reached later. In the example, pre-cooling the racks from 24 to 21°C raises the headroom from 8 to 11 K and the duration at 0.4 K per minute from 20 to 27.5 minutes. The sequence is

$$\text{pre-cool} \longrightarrow \text{store thermal margin} \longrightarrow \text{curtail later},$$

the thermal analog of charging a battery before discharging it.

The analogy extends to the costs. Pre-cooling spends energy before the event, at lower chilled-water temperatures where the chillers are less efficient (Chapter 7), so the round trip has a loss. It is bounded below by the lowest allowed inlet temperature and by the dew point of the room air, below which surfaces condense moisture. And its store leaks: the thermal mass warms back toward its normal state if the event is late, at a rate set by the same time constant that made the storage possible.

Figure 21.3 plots the rack temperature under reduced cooling for four illustrative conditions. The horizontal axis is minutes from the start of the curtailment; the vertical axis is the rack temperature. The dashed red line is the limit $T_{\max} = 32$°C. Each straight line starts at the rack's temperature when the curtailment begins and rises at the rate shown in the legend; the point where it meets the limit is its t_{limit}, marked above the line. The reference case, 24°C at 0.4 K per minute, reaches the limit at 20 minutes; pre-cooling to 21°C moves the crossing to 27.5 minutes. The other two lines anticipate the next subsection.

21.3.4 Dependence on conditions

The thermal flexibility of a facility is a function of its state and its surroundings,

$$F_{\text{thermal}} = f(T,\ P_{\text{IT}},\ T_{\text{amb}},\ \dot{Q}_{\text{avail}},\ C_{\text{th}}), \tag{21.17}$$

where T stands for the present rack and water temperatures, P_{IT} for the heat the IT load is producing, T_{amb} for the outdoor temperature that sets the plant's capacity and efficiency, $\dot{Q}_{\text{avail}}$ for the cooling capacity that remains during the curtailment, and C_{th} for the thermal mass. On a cool day the racks run well below their limit, the plant has spare capacity and recovers quickly, and the curtailment can be deep and long. On a hot afternoon with racks already near their limit it is nearly zero.

The two remaining lines of Figure 21.3 show the range. On an illustrative cool day the racks start at 22°C and, with the plant's spare capacity absorbing part of the shed heat, rise at only 0.3 K per minute: 33.3 minutes to the limit. On a hot afternoon the racks start at 30°C and, with the plant already near full output and the IT load high, rise at 0.6 K per minute: $2/0.6 = 3.3$ minutes, too short for any product of practical length. The same facility, with the same equipment, offers ten times more thermal flexibility on one day than on another, and the hot afternoon is often exactly when the grid asks for it.

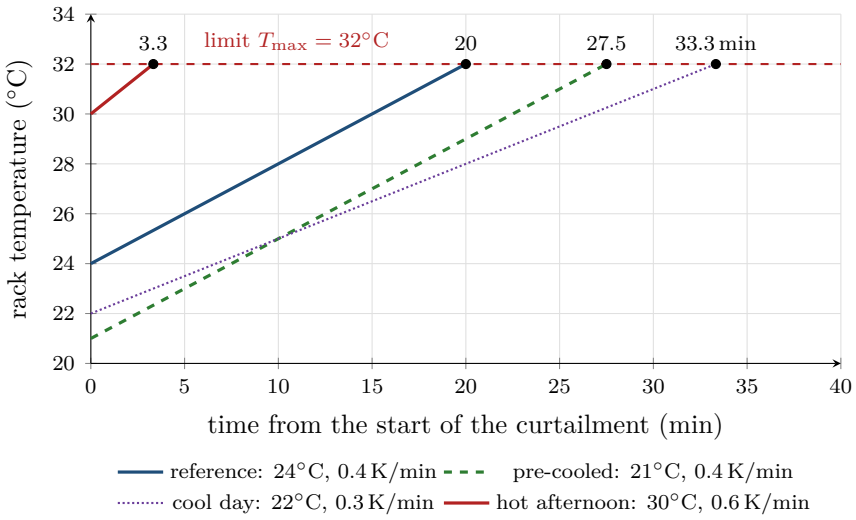


Figure 21.3: Duration of thermal flexibility, $t_{\text{limit}} = (T_{\max} - T)/(\mathrm{d}T/\mathrm{d}t)$ (equation (21.16)), for four illustrative starting temperatures and rates of rise under reduced cooling. The marks give the time to the 32°C limit in minutes. Linear extrapolation of the initial rate is conservative.

21.3.5 Backup generators

The generators can supply part of the facility's demand. A facility consuming 50 MW with its generators supplying 10 MW draws 40 MW from the grid, a 10 MW reduction that can be held as long as there is fuel. The generators are, however, not automatically attractive, for five reasons.

- *Fuel cost.* An illustrative diesel generator burns about 0.27 liters per kilowatt hour, 2700 liters for 10 MWh; at an illustrative \$1.10 per liter that is about \$300 per MWh of fuel alone, at the level of the illustrative curtailment payment of Section 21.2.6 and well above the other resources.
- *Emissions.* The same 2700 liters emit about 7 tonnes of carbon dioxide (about 2.7 kg per liter), plus nitrogen oxides and particulates at ground level near the site.
- *Equipment wear.* Every start and every running hour counts toward maintenance intervals.
- *Regulatory restrictions.* Emergency generators are usually permitted for emergencies and testing; air permits often cap non-emergency running hours per year, and running in parallel with the grid requires interconnection approval and protection.
- *Emergency readiness.* Fuel burned for a grid event is fuel not available for an outage (Chapter 4), and a generator already running for a grid event changes the transfer sequence the facility relies on.

Generators are a resource of last resort for grid flexibility: large and long-lasting, but expensive, polluting and entangled with the backup duty they exist for.

21.3.6 The portfolio and its resources

No single resource is the facility's flexibility; the facility is a portfolio. An illustrative request for 8 MW of curtailment can be met as

$$3\,\text{MW battery} + 2\,\text{MW cooling reduction} + 2\,\text{MW workload deferral} + 1\,\text{MW GPU power capping} = 8\,\text{MW},$$

each resource contributing what its own state allows. Table 21.1 compares the resources on the attributes of Section 21.1.3. Each row is a resource; the columns give how fast it responds, what limits how long it can be held, whether it produces a rebound, and what it costs or risks. The table shows why a portfolio is needed: the battery and power capping are the only very fast resources, but the battery is short and capping costs throughput; deferral and cooling are cheap but limited by deadlines and headroom; generators are long but costly. A long, deep, fast product needs several resources, each covering the part of the trajectory it is suited to.

Table 21.1: Characteristics of the flexibility resources of a data center (qualitative).

Resource	Response	Duration limited by	Rebound	Cost or penalty
battery	very fast	energy above the reserve	recharge	backup reserve, degradation
power capping	very fast	potentially long	low	computing throughput
workload deferral	moderate	job deadlines	yes	job delay
cooling reduction	fast	thermal headroom	often	temperature rise, thermal risk
generators	moderate	fuel	none	fuel, emissions, wear, permits

21.3.7 Response time and the droop law

Grid services differ in how fast they need the response. Frequency response must act within about a second; fast reserves within seconds to tens of seconds, such as the 30 s of the product of Figure 21.2; demand response within minutes, often with notice of an hour or a day. The battery responds in a fraction of a second, because its converter changes power electronically; power capping in seconds; workload deferral and migration in tens of seconds to minutes, the time to checkpoint and pause jobs; a cooling setpoint change in minutes, as the plant settles; chiller staging more slowly still, since a chiller is unloaded and stopped over many minutes; and a generator in seconds to start but minutes to synchronize with the grid and take load. Ramp capability is therefore an attribute of each resource, and it decides which products the resource can serve.

Figure 21.4 places the resources on two logarithmic axes: the horizontal axis is the response time in seconds, the vertical axis the duration for which the response can be sustained, in minutes. Each rectangle is the illustrative range of one resource. The two vertical dashed lines mark the response times of frequency response (1 s) and of the 30 s product. Only the battery lies to the left of the first line, with power capping just to its right; both lie well to the left of the second, and workload deferral, cooling and the generators lie at or to the right of it. Duration rises toward the top: the generators are the only resource that lasts a day or more. The figure also shows how a portfolio is sequenced. A fast resource bridges a slow one: the battery provides the first minutes of an 8 MW curtailment while workload pauses and the cooling plant unloads, then hands over. Bridging is cheap in energy, 5 MW for 3 minutes is 0.25 MWh, but it requires the battery to hold that energy above its reserve.

Frequency response. When a large generator trips, the grid frequency falls, and resources that respond automatically to the frequency arrest the fall. A battery can be controlled to raise its discharge in proportion to the deviation,

$$\Delta P_{\text{batt}} = -K_f\,(f - f_0), \tag{21.18}$$

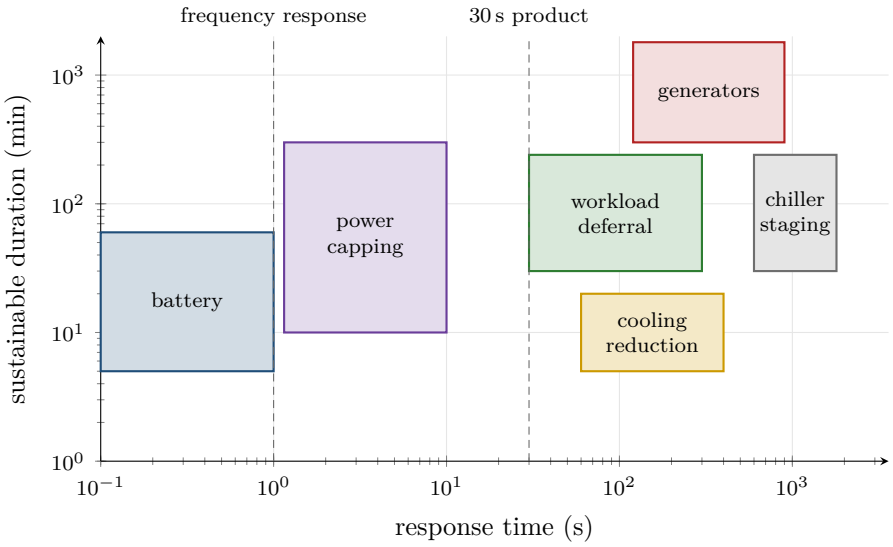


Figure 21.4: Response time and sustainable duration of the flexibility resources, illustrative ranges on logarithmic axes. Dashed lines: the response times required by frequency response and by the 30 s product of Figure 21.2. Fast resources bridge slow ones.

where f is the measured frequency, f_0 its nominal value and K_f the gain in megawatts per hertz, usually with a small deadband around f_0 in which nothing happens. This is a droop law, the same proportional control that governs conventional generators. With an illustrative $K_f = 20\,\text{MW}$ per hertz on a 60 Hz system, a fall to 59.9 Hz raises the battery's discharge by 2 MW; a fall to 59.75 Hz calls for 5 MW, the battery's grid-service limit, beyond which the response saturates. The law works the same way on a 50 Hz system. Controllable loads can follow it too: a power cap tightened in proportion to the frequency deviation reduces consumption with the same sign. Only resources at the first dashed line of Figure 21.4 can take part: the battery, and power capping where its firmware acts within about a second.

21.4 Baselines and forecasts of flexibility

A grid request names a period and a quantity; whether the facility can accept it, and whether it then delivered, both depend on quantities that cannot be measured: the demand it would have drawn without the request, and the flexibility it will have at a future hour. This section treats both as predictions with uncertainty.

21.4.1 Demand response and safe flexibility

Consider a request: "reduce demand by 5 MW from 17:00 to 18:00". Deciding whether it can be done safely requires predictions, made in advance, of the workload during that hour (how much is deferrable and how much slack the deadlines leave), of the cooling load and the thermal trajectory (how warm the racks will be at 17:00 and how fast they will rise), of the battery's state of charge at 17:00, and of the backup requirement (whether an elevated outage risk will raise the reserve). The question is not whether the equipment can physically reduce demand by 5 MW, a technical flexibility F_{tech}, but whether it can do so without violating any thermal, electrical, reserve or service constraint, a safe flexibility F_{safe}. Always $F_{\text{safe}} \le F_{\text{tech}}$: shutting down half the servers is technically possible and almost never safe for the service.

21.4.2 Baselines

A claimed curtailment of 5 MW is a reduction relative to the baseline, and the baseline is the demand that would have occurred without the event. It cannot be measured, because the event did occur; it can only be predicted, from the facility's history, the workload schedule and the weather. Suppose the facility expected 50 MW and measured 45 MW during the event. The apparent response is 5 MW. If the facility would in fact have drawn only 47 MW during that hour, because a batch of jobs finished early or the weather was cooler than forecast, then the real response was $47 - 45 = 2$ MW, and three of the claimed five megawatts were a forecast error.

Demand response therefore requires counterfactual forecasting. The three kinds of statement of this book appear in their sharpest form here: the measured grid power during the event is a meter reading; the baseline is inferred, a prediction of a quantity that never existed; and the delivered flexibility, their difference,

$$\Delta P_{\text{flex}} = P^{\text{base}} - P_{\text{grid}}, \tag{21.19}$$

is calculated from one measured and one inferred quantity and inherits the uncertainty of the inferred one. Settlement rules in demand-response programs fix a baseline method, such as the average of the same hours on several recent similar days, precisely because the counterfactual has to be agreed before it can be paid for.

Figure 21.5 draws the example. The horizontal axis is the time of day; the vertical axis is net grid power. The dashed blue line is the forecast baseline of 50 MW; the dark band around it is one standard deviation of the forecast, ±2 MW, and the light band two standard deviations. The solid black line is the measured demand: 50 MW before the event, 45 MW during it, a short rebound above 50 MW after it. The dotted red segment is one possible true baseline during the event, 47 MW. The two vertical arrows are the apparent and the real response. The figure shows that the difference between the two is well inside the forecast band: a true baseline of 47 MW is only 1.5 standard deviations below the forecast and entirely plausible.

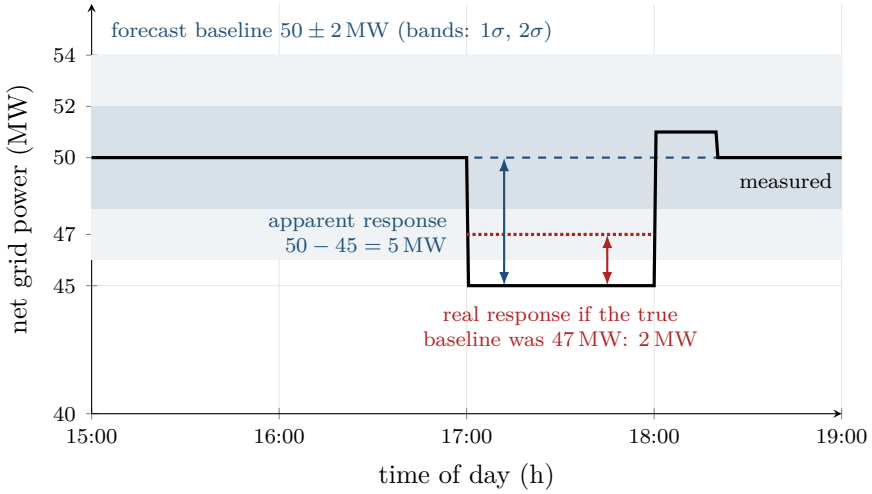


Figure 21.5: The baseline is a counterfactual. Forecast baseline 50 MW with an illustrative standard deviation of 2 MW; measured demand 45 MW during a 17:00 to 18:00 event. The claimed 5 MW response becomes 2 MW if the true baseline was 47 MW, a value 1.5 standard deviations below the forecast.

Principle 21.2. *Delivered flexibility is the difference between a measured demand and an inferred counterfactual; it is uncertain by at least the uncertainty of the baseline forecast.*

21.4.3 Baseline uncertainty

If the baseline forecast is $50 \pm 2\,\mathrm{MW}$, read as a normal distribution with mean 50 MW and standard deviation 2 MW, and the measured demand is 45 MW with negligible meter error, then by equation (21.19) the delivered flexibility is itself a distribution,

$$p(\Delta P_{\text{flex}} \mid \text{baseline uncertainty}) = \mathcal{N}(50 - 45,\ 2^2) = \mathcal{N}(5\,\mathrm{MW},\ (2\,\mathrm{MW})^2). \tag{21.20}$$

Its median is the apparent 5 MW, but with probability 0.5 the response was less. The probability that at least 3 MW was delivered is $1 - \Phi\big((3-5)/2\big) = \Phi(1) = 0.84$; the probability that less than 2 MW was delivered, the case of the true baseline below 47 MW, is $\Phi(-1.5) = 0.067$. A 90 % interval for the delivered flexibility is $5 \pm 1.645 \times 2$, from 1.7 to 8.3 MW.

Two consequences follow. A single event cannot establish whether a facility delivered 5 MW to better than about ± 3 MW; a record of many events, whose baseline errors partly average out, can establish the average response far more precisely. And the baseline uncertainty is set by how predictable the facility's demand is: a facility whose IT load is steady has a narrow band and its flexibility is verifiable; one whose batch load swings by several megawatts from hour to hour has a wide band, and its delivered flexibility is uncertain even when the meter is perfect.

21.4.4 Forecasting flexibility

An offer of flexibility for tomorrow afternoon requires forecasts of every input to the feasible range of equation (21.2) at that hour: $P_{\text{IT}}(t+h)$, $P_{\text{cool}}(t+h)$, $T_{\text{amb}}(t+h)$ and $\text{SOC}(t+h)$ for a horizon h, together with the job queue and its deadlines. From them the flexibility itself is forecast,

$$F(t+h) = g\big(P_{\text{IT}}(t+h),\ P_{\text{cool}}(t+h),\ T_{\text{amb}}(t+h),\ \text{SOC}(t+h),\ \ldots\big), \tag{21.21}$$

where g is the computation of Sections 21.2 and 21.3: the battery's energy above its reserve, the thermal headroom and rate of rise, the deferrable work and its slack, the capping allowed. Flexibility is itself a forecastable quantity, and its forecast inherits the uncertainty of every input.

An illustrative case shows how much it can change within a day. At 10:00 the facility's safe downward flexibility is $F^{\downarrow}_{\text{safe}} = 8\,\mathrm{MW}$, the portfolio of Section 21.3.6. The forecast for 17:00 is 3 MW, because three conditions change: the IT demand is higher, so less work is deferrable; the weather is hotter, so the plant has no spare capacity and the rate of rise is high; and the thermal headroom is smaller, because the racks run warmer. Table 21.2 breaks the two numbers down by resource; its first two numerical columns are this example, and its third belongs to Section 21.5.2. The cooling contribution vanishes and the deferral shrinks, while the battery is lower because part of its energy was used for a morning peak. The 5 MW request of Section 21.4.1 for 17:00 to 18:00 is technically within the facility's means and, by this forecast, not safe.

21.5 Committing flexibility under uncertainty

A forecast of flexibility is a distribution, and a commitment to the grid is a single number. This section chooses that number: as a quantile of the forecast, as the result of a chance constraint, and as an envelope over durations, and it adds the rebound that every commitment carries.

21.5.1 Firm flexibility

Suppose that for an illustrative afternoon window, 14:00 to 15:00 on the same day, the forecast of safe downward flexibility gives $\Pr(F \geq 5\,\mathrm{MW}) = 0.70$, while only 3.5 MW can be guaranteed

Table 21.2: Safe downward flexibility by resource, illustrative: now at 10:00, forecast for 17:00, and on the day of an overcommitted event.

Resource	10:00 (MW)	17:00 forecast (MW)	event day (MW)	What changed
battery	3.0	1.5	2.5	state of charge lower after the morning; lower on the event day
cooling reduction	2.0	0.0	0.0	hot afternoon; on the event day one chiller unavailable
workload deferral	2.0	0.5	0.5	higher IT demand, deadlines inside the window
GPU power capping	1.0	1.0	1.0	unchanged
total	8.0	3.0	4.0	

at 95 % confidence. The second number is the firm offer,

$$\Pr(F \geq F_{95}) = 0.95, \qquad F_{95} = 3.5\,\mathrm{MW}, \tag{21.22}$$

the quantity the facility can commit with a 5 % probability of not being able to deliver it. It is the same construction as the quantile capacity C_{95} of Chapter 17, applied to flexibility instead of capacity: a risk-constrained quantity rather than an expected one.

The two statements fix a normal forecast completely. Writing $F \sim \mathcal{N}(\mu, \sigma^2)$, the first says $5 = \mu - 0.524\,\sigma$ and the second $3.5 = \mu - 1.645\,\sigma$; subtracting, $1.5 = 1.121\,\sigma$, so

$$\sigma = 1.34\,\mathrm{MW}, \qquad \mu = 5.70\,\mathrm{MW}. \tag{21.23}$$

The expected flexibility in that window is 5.7 MW, the firm flexibility 3.5 MW, and the gap of 2.2 MW is the price of a 95 % guarantee under this uncertainty. A sharper forecast, with a smaller σ, would raise F_{95} toward μ: every reduction of the forecast's uncertainty, from better estimates of the state of charge, the thermal parameters or the job queue, becomes committable megawatts.

Figure 21.6 plots this forecast. The horizontal axis is the safe flexibility F in megawatts; the vertical axis is its probability density. The curve is the normal density of equation (21.23). The green area to the right of 5 MW holds 70 % of the probability; the red area to the left of F_{95} holds 5 %. The vertical lines mark F_{99}, F_{95}, 5 MW and a commitment of 8 MW that the next subsection discusses. The figure shows how far into the left tail a high-confidence commitment must reach: the 95 % and 99 % firm values lie well below the peak.

21.5.2 Overcommitment

Suppose the facility commits 8 MW, the safe flexibility of a cool morning, for a later window. On the day of the event three things differ from the plan: the workload is higher than expected, so most jobs have no slack; one chiller is unavailable, so the remaining plant runs at full output and no cooling can be shed; and the battery's state of charge is lower than planned. The last column of Table 21.2 gives the result: the safe flexibility is 4 MW, half the commitment.

The facility then has three choices, all bad. It can fail to deliver and pay the program's penalty. It can deliver by compromising computing: deferring time-sensitive work or capping deeper than its service agreements allow. Or it can deliver by reducing resilience: discharging the battery below its reserve or shedding cooling into the thermal margin. The last two convert

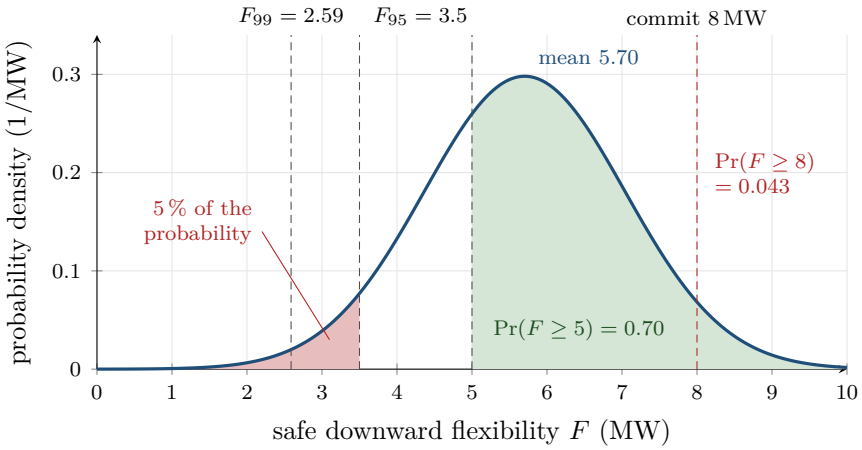


Figure 21.6: Forecast of safe downward flexibility for an illustrative window, $F \sim \mathcal{N}(5.70, 1.34^2)$ MW (equation (21.23)), fixed by $\Pr(F \geq 5) = 0.70$ and $F_{95} = 3.5$ MW. Red: the 5 % below F_{95}; green: the 70 % above 5 MW. The 99 % firm value is $F_{99} = 2.59$ MW; an 8 MW commitment is safe with probability 0.043.

a commercial problem into a service or safety problem, which the facility's primary mission forbids. The uncertainty must be accounted for before the commitment, not discovered during the event. Against the forecast of Figure 21.6, an 8 MW commitment is 1.72 standard deviations above the mean and is safe with probability $1 - \Phi(1.72) = 0.043$: the overcommitment was predictable from the distribution, not only from hindsight.

21.5.3 The chance-constrained offer

The firm offer generalizes to a chance constraint over all the facility's constraints together,

$$F^* = \max \left\{ F : \Pr\big(\text{all data center constraints remain satisfied} \mid F\big) \geq 0.99 \right\}, \tag{21.24}$$

the largest commitment for which the probability that every thermal, electrical, reserve and service constraint holds during and after the event is at least 0.99. The result is a reliable flexibility envelope rather than a single forecast. When all the uncertainty is summarized in one forecast of the safe flexibility, as in Figure 21.6, delivering F is safe exactly when $F_{\text{safe}} \geq F$, and F^* is the 1 % quantile of the forecast: $F^* = F_{99} = \mu - 2.326\,\sigma = 5.70 - 3.11 = 2.59$ MW. Going from 95 % to 99 % confidence costs 0.9 MW of offer.

The "all" in equation (21.24) matters. A joint chance constraint over several constraints is stronger than separate ones: if the thermal, electrical and reserve constraints each hold with probability 0.99, all three hold with a probability that may be as low as 0.97. A common practical approach divides the tolerance among them, for example 0.004 for thermal, 0.003 for electrical and 0.003 for the reserve, so that the union of their failures has probability at most 0.01 (Chapter 20 develops chance constraints in general).

21.5.4 The flexibility envelope

Because each resource is limited differently in duration, the safe flexibility depends on the length of the event and on when it starts. The complete statement is an envelope $F(\Delta P, \tau, t_{\text{start}})$: how much, for how long, starting when. For an illustrative facility at a given start time it might read: 8 MW for 5 minutes, 5 MW for 30 minutes, 2 MW for 2 hours. The envelope's boundary is the sum of the resources' own limits for a sustained event of duration τ,

$$\overline{F}(\tau) = \sum_{k} \min\left(P_k^{\lim},\ \frac{E_k}{\tau} \right), \tag{21.25}$$

where $P_k^{\lim}$ is the power limit of resource k and E_k the energy it can shift (infinite for power capping, which depletes nothing).

Figure 21.7 draws equation (21.25) for that illustrative facility. The horizontal axis is the event duration on a logarithmic scale, from 5 minutes to 3 hours; the vertical axis is the sustained curtailment. The colored bands are stacked: from the bottom, power capping (1 MW at any duration), workload deferral (2 MW up to 0.5 MWh), cooling reduction (2 MW up to 0.67 MWh, the 2 MW for 20 minutes of Section 21.3.2), and the battery (3 MW up to 0.83 MWh above its reserve, at a state of charge of about 0.68). The top edge is the envelope. Up to 15 minutes every resource runs at its power limit and the envelope is flat at 8 MW; beyond that the energy-limited resources drop out one after another, the envelope passes through 5 MW at 30 minutes and 2 MW at 2 hours, and at long durations only power capping remains. Any commitment on or below the top edge is feasible at that start time; one above it is not. The product of Figure 21.2, 5 MW for 30 minutes, lies exactly on the envelope: it uses everything the facility has at that duration.

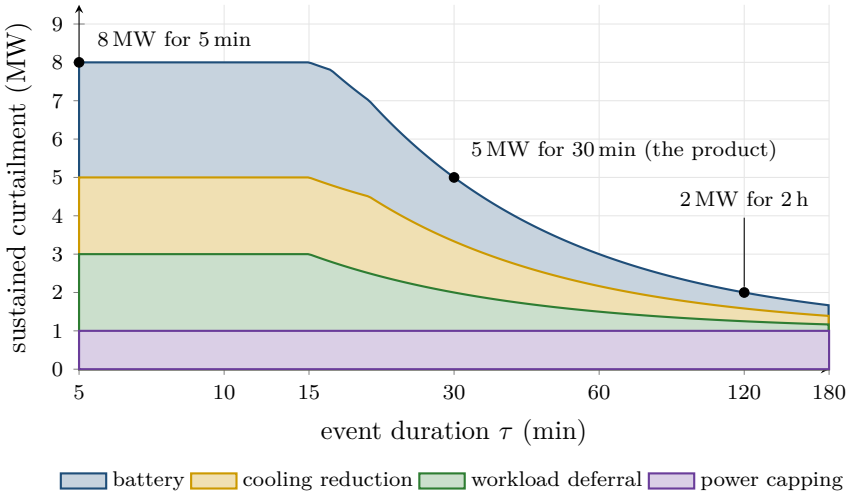


Figure 21.7: An illustrative flexibility envelope, $\overline{F}(\tau) = \sum_k \min(P_k^{\lim}, E_k/\tau)$ (equation (21.25)), with τ in minutes: capping 1 MW (unlimited energy), deferral $\min(2, 30/\tau)$, cooling $\min(2, 40/\tau)$, battery $\min(3, 50/\tau)$, the energies in megawatt minutes. Stacked bands; the top edge is the envelope.

The envelope also depends on t_{start}. The same facility evaluated for 17:00 on a hot afternoon has no cooling band and a thinner deferral band, and its whole envelope sits lower; the forecast of Section 21.4.4 is one point of that later envelope. An offer to the grid is a point under the envelope of its own start time, and a firm offer is a point under the corresponding quantile of the envelope.

21.5.5 Rebound constraints

Most flexibility is temporal shifting rather than energy reduction. Deferring 1 MWh of computing does not remove it: the energy is consumed later. Battery discharge must be followed by recharge, of slightly more energy than was discharged because of conversion losses. Cooling curtailment leaves heat in the thermal mass that the plant must remove afterward, with extra cooling. Over a recovery horizon τ_r after an event of duration τ starting at t_s, the change in energy is therefore close to zero,

$$\int_{t_s}^{t_s+\tau+\tau_r} \Delta P(t)\,\mathrm{d}t \;\approx\; 0, \tag{21.26}$$

up to a positive remainder (the battery's round-trip loss, the extra energy of cooling at a worse efficiency during recovery) and a negative remainder (work genuinely shed, as when a capped job is not completed rather than slowed).

The product of Figure 21.2 shows what the balance constrains. Its curtailment removes 2.5 MWh; its permitted rebound, 2 MW for 20 minutes, returns only 0.67 MWh. The remaining 1.83 MWh must either be work that is not consumed at all, which in a data center means power capping with throughput lost rather than delayed, or be returned outside the product's rebound window, spread over the following hours at a rate the grid does not object to. If the 5 MW were provided as 3 MW of battery, 1 MW of deferral and 1 MW of capping, the battery would need 1.5 MWh/0.9 = 1.67 MWh of recharge and the deferred work 0.5 MWh, in all 2.17 MWh, more than three times the window; the battery would have to recharge at about 0.5 MW for three hours. A rebound limit is a constraint on how the portfolio is dispatched and on how fast the energy is returned, and it has to be planned together with the curtailment. A badly placed rebound can also create the problem the event was meant to relieve: if every facility in a region recharges and restarts its deferred work at 18:00, the peak moves to 18:00.

21.6 Reserve, grid signals and reporting

The last group of ideas sets the priority between grid service and resilience, places the data center among other aggregated resources, adds price and carbon as signals that drive flexibility, and ends with what a facility should report and the compact formulation of the whole problem.

21.6.1 The reliability reserve has priority

Resilience comes before grid revenue. Take the battery of Section 21.2.2 at full charge, 10 MWh usable, and an event of one hour. Without a reserve the battery could offer 10 MW for the hour. If resilience requires a reserve of 6 MWh, the $\mathrm{SOC}_{\mathrm{reserve}} = 0.6$ of equation (21.7), only 4 MWh is available, and the offer is 4 MW. The reserve is not fixed. If the risk of an outage rises, a storm warning, a grid alert, a planned switching operation upstream, the reserve required for the same probability of a failed ride-through rises, and the flexibility falls. The reserve should therefore be dynamic and risk-based, recomputed from the current risk, rather than a constant set once at commissioning.

21.6.2 A risk-based reserve

The reserve can be computed from equation (21.8) once its uncertain inputs have probabilities. Consider an illustrative model of the next hour. A utility outage occurs with probability p_o. If it does, the generators start at the first attempt with probability $1-q = 0.95$, and the battery must then carry the critical load for $t_1 = 2$ minutes (start, transfer and margin); otherwise it must carry it for $t_2 = 8$ minutes (a second attempt or a manual start). The usable energy at full charge is uncertain because the state of health is estimated, not measured: $E_{\mathrm{usable}} \sim \mathcal{N}(10, \sigma_E^2)$ MWh with $\sigma_E = 1$ MWh. At a state of charge s the runtime is $t_{\mathrm{backup}} = 60\, s\, E_{\mathrm{usable}}/40$ minutes. The probability of a failed ride-through in the next hour is

$$\Pr(\text{backup failure} \mid s) = p_o\Big[(1-q)\,\Pr\big(t_{\mathrm{backup}} < t_1\big) + q\,\Pr\big(t_{\mathrm{backup}} < t_2\big)\Big], \tag{21.27}$$

and the reserve is the smallest s for which it is at most $\epsilon_B = 10^{-6}$ per hour. This is a probability of backup failure given the state of charge, the state of health and the generator state, $\Pr(\text{backup failure} \mid \mathrm{SOC}, \mathrm{SOH}, \text{generator state})$; the minimum safe $\mathrm{SOC}_{\mathrm{reserve}}$ follows from it, and the energy above it is what grid participation may use.

For a normal hour with $p_o = 10^{-4}$, the bracket may be as large as 0.01, which allows $\Pr(t_{\mathrm{backup}} < 8\,\mathrm{min}) \le 0.2$; the runtime at state of charge s has mean $15s$ and standard deviation

1.5s minutes, so s must satisfy $15s - 0.84 \times 1.5s \geq 8$, which gives $\mathrm{SOC}_{\mathrm{reserve}} = 0.58$. In an hour with a storm warning, $p_o = 10^{-2}$, the bracket may be at most 10^{-4}, which allows $\Pr(t_{\mathrm{backup}} < 8\,\mathrm{min}) \leq 0.002$, that is $15s - 2.88 \times 1.5s \geq 8$, and the reserve rises to 0.75. From full charge, the one-hour offer is 4.2 MW in the normal hour and 2.5 MW in the storm hour. The fixed rule of 0.6 is close to the risk-based reserve in normal conditions and badly short of it in the storm: at SOC = 0.6 the failure probability in the storm hour is 6.7×10^{-5}, about 67 times the tolerance.

Figure 21.8 plots the risk-based reserve against the outage probability. The horizontal axis is p_o on a logarithmic scale; the vertical axis is the minimum safe state of charge. The solid curve is the case above, $\sigma_E = 1$ MWh; the dashed curve halves the uncertainty of the usable energy, $\sigma_E = 0.5$ MWh, as a better estimate of the state of health would. The horizontal dotted line is the fixed rule of 0.6. Both curves rise with the outage probability, and they separate as it rises: in the storm hour the better estimate of the battery's health lowers the reserve from 0.75 to 0.62, freeing 1.3 MWh for the grid without any change in the risk. The reserve links three subjects of this book: diagnosis and estimation, which set σ_E (Chapter 14); reliability, which sets q and ϵ_B (Chapter 12); and grid participation, which uses what is left.

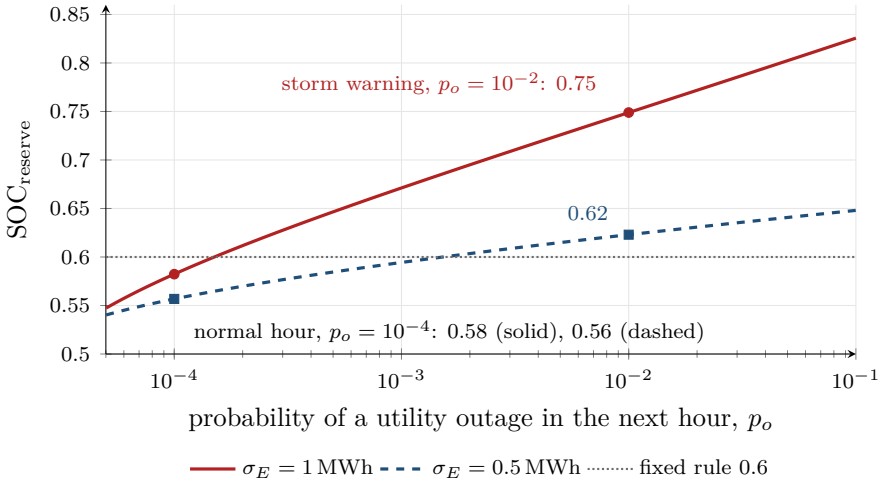


Figure 21.8: Risk-based battery reserve, the smallest state of charge for which the failure probability of equation (21.27) is at most 10^{-6} per hour. Illustrative model: 10 MWh usable (mean), 40 MW critical load, generators fail the first start with probability 0.05, required runtime 2 or 8 minutes. Computed by `scripts/ch21_data.py`.

Principle 21.3. *The backup reserve has priority over grid service and is not a constant: it is the smallest state of charge whose probability of a failed ride-through is acceptable, and it rises with the risk of an outage and with the uncertainty of the battery's health.*

21.6.3 A portfolio behind one meter and the aggregation analogy

Seen from the grid, a data center is a portfolio of distributed energy resources behind one meter (Figure 21.1): a battery, flexible load, thermal storage in the form of thermal mass and chilled water, and backup generation. Its primary mission is computing, and grid flexibility is subordinate to service and resilience: every resource in the portfolio has a first duty that is not to the grid.

The structure is the same as that of an aggregator of distributed energy resources, a virtual power plant. An aggregator combines thousands of household thermostats, electric vehicle chargers, home batteries and water heaters, each with its own owner and constraints (comfort,

a car that must be charged by morning, hot water at breakfast), and offers their combined flexibility to the grid. A data center combines IT load, a battery, cooling and thermal inertia, each with its own constraint (deadlines, the backup reserve, the thermal limits). Both follow the same cycle: forecast the flexibility, commit an offer, dispatch the resources, verify the delivery against a baseline. The differences are of degree: the data center's resources are few, large and under one operator's control, the aggregator's many, small and under the owners' control; the data center's constraints are hard (a failed ride-through or an overheated rack), the aggregator's mostly soft (a degree of comfort).

Figure 21.9 draws the cycle. The four boxes in the middle row are its stages, from left to right. The box above the first stage holds the inputs: estimates of the present state (state of charge, state of health, temperatures) and forecasts (IT load, weather, prices). The box above the second and third stages holds the constraints that every commitment and dispatch must respect. The arrow below the row returns the verified delivery to the next forecast, where it improves the baseline model and the estimates of the thermal and battery parameters. The two notes at the bottom list the resources an aggregator and a data center bring to the same cycle.

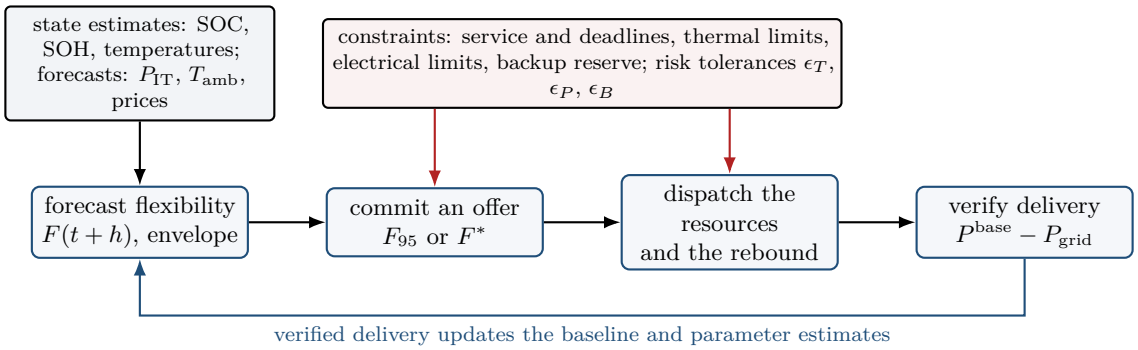


Figure 21.9: The flexibility cycle shared by an aggregator of distributed energy resources and a data center: forecast, commit, dispatch, verify, with state estimates and forecasts as inputs, the facility's constraints on every commitment and dispatch, and the verified delivery fed back.

21.6.4 Price and carbon signals

Grid signals can influence computing directly. Suppose the price of electricity spikes at 17:00, to an illustrative \$1000 per MWh against \$80 in the morning. The facility can respond by deferring training jobs, moving workloads to another site, discharging the battery within its reserve, and altering its cooling strategy, for example pre-cooling in the cheaper hours before the spike. A facility that responds this way is no longer a passive load that draws whatever its computers demand; it is a cyber-physical energy system, in which computational decisions and physical energy flows are optimized together (Chapter 20 works a batch-job schedule against an hourly price).

The same structure serves a carbon objective. The carbon intensity of grid electricity, $\mathrm{CI}(t)$, varies through the day with the mix of generation: low when solar or wind output is high, high when fossil units set the margin. A carbon-aware schedule solves

$$\min \sum_t \mathrm{CI}(t)\, P_{\mathrm{grid}}(t)\, \Delta t \quad \text{subject to the deadlines and the physical limits,} \tag{21.28}$$

with the emissions of any on-site generation added if the generators run. Moving the illustrative 2 MWh job j of Section 21.2.4 from an evening hour with $\mathrm{CI} = 0.45\,\mathrm{t/MWh}$ to a midday hour

with 0.25 t/MWh, if its deadline allows, saves $2 \times (0.45 - 0.25) = 0.4$ tonnes of carbon dioxide; the same move against prices would be valued at the price difference. The two objectives often agree, since high prices and high carbon intensity both mark hours when the least efficient fossil units run, but not always, and a facility that pursues both needs a weighted objective of the kind Chapter 20 discusses.

21.6.5 Geographic shifting

A company with data centers in several regions can move flexible work between them. If region A is congested, expensive and carbon-intensive while region B has spare capacity, flexible workloads move from A to B: grid-aware geographic dispatch. With illustrative numbers, moving 3 MW of batch computing for four hours from a region at \$250 per MWh and 0.6 t/MWh to one at \$60 per MWh and 0.3 t/MWh shifts 12 MWh, saves $12 \times 190 = \$2280$ and $12 \times 0.3 = 3.6$ tonnes of carbon dioxide, and relieves the congested grid of 3 MW for four hours. The limits are those of Chapter 20: the data must be where the work runs, latency-sensitive services cannot move, and the receiving site must have the power, cooling and thermal margin to absorb the work. For the grid, geographic shifting is a curtailment in one place and an increase in another, and it helps only when the two places are on different sides of a constraint.

21.6.6 A practical flexibility report

The ideas of the chapter come together in what a facility reports to its operators and to the grid at each moment. Table 21.3 shows an illustrative report. Each row is one statement: the current baseline, the safe downward flexibility at two durations (two points of the envelope), the firm flexibility at 95 % confidence for an hour, its composition by resource, the expected rebound, and the constraint that limits the offer. The composition sums to the firm value: $1.5 + 1.0 + 0.7 = 3.2$ MW. The rebound line gives the part of the recovery the facility expects in the period right after the event; the rest of the battery's recharge can be spread over the following hours. The last line is as important as the numbers: it tells the operator where an additional measurement, a pre-cooling action or a maintenance decision would raise the offer.

Table 21.3: An illustrative flexibility report.

Item	Value
current baseline	52 MW
safe downward flexibility	6.5 MW for 15 min; 4.0 MW for 60 min
firm flexibility (95 %)	3.2 MW for one hour
composition of the firm value	battery 1.5 MW, workload deferral 1.0 MW, cooling reduction 0.7 MW
expected rebound	1.8 MW for 40 min after the event
primary constraint	thermal margin in hall C

21.6.7 The compact formulation

The question the report answers is how much flexibility can safely be committed now, given uncertain future load, weather, equipment state and resilience requirements. Answering it requires forecasting (the load, the weather, the baseline), thermal and electrical physics (the rates of rise, the power balances), state estimation (state of charge, state of health, temperatures, thermal parameters), contingency risk (the outage and generator-start probabilities) and optimization (the dispatch of the portfolio). The output is a risk-constrained flexibility forecast,

not just a demand-response forecast. In compact form, at time t,

$$\begin{aligned} \max\ & F_t \\ \text{subject to}\ & \Pr(T_i > T_{\max}) < \epsilon_T \quad \text{for every rack } i, \\ & \Pr(P_{\text{facility}} > P_{\text{lim}}) < \epsilon_P, \\ & \Pr(t_{\text{backup}} < t_{\text{required}}) < \epsilon_B, \\ & W_{\text{served}} \geq W_{\text{required}}, \end{aligned} \tag{21.29}$$

where T_i is the temperature of rack i during and after the event, P_{lim} the electrical capacity of the facility's supply path (transformers, switchgear, UPS), W_{served} the computing work delivered over the horizon and W_{required} the work the service agreements and deadlines demand. The probabilities are taken over the forecast uncertainty of the inputs and the estimation uncertainty of the states and parameters, and the tolerances ϵ_T, ϵ_P and ϵ_B are set by the consequence of each violation (Chapter 20). Each line is one of the chapter's resources seen as a constraint: the thermal line limits cooling curtailment, the electrical line limits upward flexibility and rebound, the reserve line limits the battery, and the service line limits deferral and capping.

21.7 What to retain

Definition 21.1 (Data center flexibility)**.** A data center is not only a large load but a flexible, multi-resource energy system, constrained by computing service, thermal safety, electrical limits, the battery's backup reserve and equipment reliability. Its flexibility is not how many megawatts it can shed but how many megawatts it can reliably deliver, for how long, under current and forecast conditions, without compromising its primary mission.

Table 21.4 lists the chapter's objects and the subsection that defines each.

21.8 Exercises

Exercise 21.1. A data center battery has $E_{\text{usable}} = 12\,\text{MWh}$ and supports a critical load of 45 MW; resilience requires 10 minutes of runtime. (a) Compute $\text{SOC}_{\text{reserve}}$ and the runtime at full charge. (b) With a grid-service limit of 6 MW, compute the battery's downward flexibility for a 30-minute event at $\text{SOC} = 0.95$ and at $\text{SOC} = 0.75$, using equation (21.9). (c) Repeat (a) and (b) after the battery has aged to $\text{SOH} = 0.85$, keeping the 10-minute requirement.

Exercise 21.2. Rack inlets run at 25°C with a limit of 32°C. With cooling power reduced by 1.5 MW, they rise at 0.35 K per minute. (a) Compute t_{limit} and the curtailment energy in megawatt hours. (b) Repeat with a 1 K safety margin kept for recovery. (c) The racks are pre-cooled to 22°C before the event; compute the new duration with the margin. (d) If the rate of rise is proportional to the cooling power shed, how long could 1 MW be held from 25°C with the margin?

Exercise 21.3. A facility's baseline forecast for an event hour is normal with mean 48 MW and standard deviation 1.5 MW, and its measured demand during the event is 44 MW. (a) Write the distribution of the delivered flexibility. (b) Compute the probability that at least 3 MW was delivered and a 90 % interval. (c) The facility had committed 4 MW. What is the probability that it delivered its commitment?

Exercise 21.4. The forecast of safe downward flexibility for tomorrow's 17:00 to 18:00 window is normal with mean 6 MW and standard deviation 1.5 MW. (a) Compute F_{95} and F_{99} and $\Pr(F \geq 5\,\text{MW})$. (b) The facility is asked for 7 MW; compute the probability that this commitment is safe. (c) A better estimate of the battery's state of health would reduce the

Table 21.4: The objects of grid interaction and flexible load, and where each is defined.

Object	Expression	Role	Defined in
Net grid power	equation (21.1)	what the grid sees at the meter	21.1.1
Directional flexibility	$F^{\downarrow} = P^{\text{base}} - \underline{P}$, $F^{\uparrow} = \overline{P} - P^{\text{base}}$	distances to the feasible range	21.1.2
Flexibility trajectory	$\Delta P(t) = P_{\text{grid}}(t) - P^{\text{base}}(t)$	curtailment and rebound together	21.1.4
Backup reserve	$\Pr(t_{\text{backup}} < t_{\text{required}}) < \epsilon_B$	priority claim on the battery	21.2.2
Battery flexibility	equation (21.9)	state-dependent, above the reserve	21.2.3
Deadline constraint	$\int P_j \, \mathrm{d}t = E_j$ by d_j	conservation of deferred work	21.2.4
Service cost	C_{flex}, equation (21.13)	merit order of resources	21.2.6
Thermal duration	$t_{\text{limit}} = (T_{\max} - T)/(\mathrm{d}T/\mathrm{d}t)$	how long a cooling curtailment lasts	21.3.2
Droop law	$\Delta P_{\text{batt}} = -K_f(f - f_0)$	automatic frequency response	21.3.7
Delivered flexibility	$\Delta P_{\text{flex}} = P^{\text{base}} - P_{\text{grid}}$	measured minus inferred; a distribution	21.4.2
Firm flexibility	$\Pr(F \geq F_{95}) = 0.95$	committable quantity	21.5.1
Chance-constrained offer	F^*, equation (21.24)	joint safety at 0.99	21.5.3
Envelope	$\overline{F}(\tau) = \sum_k \min(P_k^{\lim}, E_k/\tau)$	how much, how long, starting when	21.5.4
Rebound balance	equation (21.26)	shifting rather than reduction	21.5.5
Risk-based reserve	equation (21.27)	reserve as a function of outage risk	21.6.2
Carbon objective	equation (21.28)	schedule against CI(t)	21.6.4
Compact formulation	equation (21.29)	risk-constrained flexibility	21.6.7

standard deviation to 1.0 MW without changing the mean. By how much does it raise F_{95}? (d) If the facility's forecast also gives $\Pr(\mathrm{SOC}_{\mathrm{reserve}} > 0.7) = 0.3$ because of a storm warning, explain in one sentence why the forecast's mean and standard deviation should be recomputed rather than the offer lowered by a fixed amount.

21.9 Connection to the TERRA framework

Grid flexibility uses every layer the framework provides: the conservation networks for the power and heat balances, forward simulation for durations and rebound, and posterior inference for the states, the baseline and the risk.

- **The meter is a conservation node.** Equation (21.1) is the power balance of the node where the facility bus meets the grid. Its edges carry P_{IT}, P_{cool}, P_{aux}, P_{batt} and P_{gen}; each component on an edge has bounds, and the feasible range of equation (21.2) is the set of node injections those bounds allow. Figure 21.1 is that node and its edges.
- **Thermal flexibility is forward simulation.** The duration of a cooling curtailment is the time at which a simulated rack node of the thermal network, equation (21.15), crosses its bound $T_{\max}$; pre-cooling is a different initial state of the same simulation, and the rebound is the simulation continued after cooling is restored. The rate of rise depends on the effective capacitance, an inferred parameter.
- **The reserve is a probability computed from a posterior.** The battery's usable energy comes from the posterior over its state of health, and the generator's start from its reliability; the risk-based reserve of equation (21.27) is a probability that available runtime falls below required runtime, the same form as a probability that available capacity is below required capacity. A tighter posterior lowers the reserve, as Figure 21.8 shows.
- **The baseline is an inferred counterfactual.** The delivered flexibility of equation (21.19) is a measured node power minus a posterior predictive of the same node without the event; its distribution, not a point, is what verification reports.
- **The firm offer is a quantile of a predictive distribution.** F_{95}, F^{*} and the envelope are computed by propagating the posterior over states and parameters and the forecast distributions of load and weather through the constraints of equation (21.29), and reading off the commitment that keeps every constraint satisfied with the required probability.

In summary, the flexibility a data center can offer the grid is a trajectory limited by the battery's energy above a reserve that resilience sets, by the thermal headroom of its rooms, by the deadlines of its computing work and by the cost of capping it; it is verified against a counterfactual baseline and committed as a quantile of a forecast, so that what is promised can be delivered with high probability without compromising computing or resilience. The next chapter turns to the models that produce such forecasts and their fidelity: digital twins, computational fluid dynamics and the reduced models that stand between them (Chapter 22).

Chapter 22

Digital Twins, CFD and Reduced Models

22.1 Three kinds of model

Every chapter of this book has used a model of the facility: a heat balance for a rack, a mixing rule for an aisle, a capacity curve for a chiller, a filter for the present state, a posterior over causes. The models differed in how much of the physics they kept, and each was chosen, usually without comment, to be just detailed enough for the question at hand. Three modeling philosophies dominate the practice of data center engineering: computational fluid dynamics, which resolves the air and temperature fields in space; reduced-order physical models, which keep a small number of states; and digital twins, which keep a model connected to the operating facility. The question of this chapter is

how much model fidelity diagnosis, capacity assessment, risk estimation and control actually require, and which kind of model supplies it.

The answer developed below is that a more detailed model is not always more useful in operation. Detail costs computing time, it needs inputs that are rarely known, and it adds unknowns that the sensors of a real facility cannot determine. The chapter describes what each kind of model represents and what it costs, works a small reduced network through by hand, explains how a model is kept connected to the facility, treats the error that every model carries, and ends with a rule for choosing the fidelity from the decision the model serves.

Jargon: computational fluid dynamics, reduced-order model, digital twin, fidelity, boundary condition, surrogate, calibration, data assimilation, model discrepancy, multi-fidelity model

Computational fluid dynamics (CFD) is the numerical solution of the conservation equations of mass, momentum and energy of a fluid on a fine grid of cells in three dimensions. A *reduced-order model* replaces those fields by a small number of states, such as one temperature per aisle or one airflow per path. A *digital twin* is a model of a specific physical asset that is connected to the asset's operational data and updated from them repeatedly. The *fidelity* of a model is how much physical detail it resolves, in space, in time and in mechanisms. A *boundary condition* is a value a model takes as given at the edge of what it computes: a fan flow, a supply temperature, a heat load. A *surrogate*, also called an *emulator* or *response surface*, is a cheap approximation fitted to the outputs of an expensive simulator. *Calibration* is the adjustment of a model's fixed parameters so that its outputs match data. *Data assimilation* is the repeated combination of a model's prediction with new measurements to update its state and parameters, as in weather forecasting. *Model discrepancy* is the systematic difference between what a model predicts and what the real system does, even with the best parameters. A *multi-fidelity model* combines a cheap low-fidelity model with a few runs of an expensive high-fidelity model that correct it.

Key idea

CFD supplies spatial detail, a reduced model supplies speed and tractable inference, and a digital twin supplies the connection between a model and the real facility. The right fidelity is not the highest one available but the lowest one that is accurate enough for the decision, with its remaining error stated explicitly.

22.1.1 The distinction

A CFD model resolves detailed fluid and thermal fields. It divides the room into a grid of small cells, often millions of them, and computes in each cell the air velocity, the pressure and the temperature. Its output is a field: a temperature at every point of the room, an air velocity at every point, from which any local quantity, the temperature at the top of one rack's inlet or the flow through one perforated floor tile, can be read.

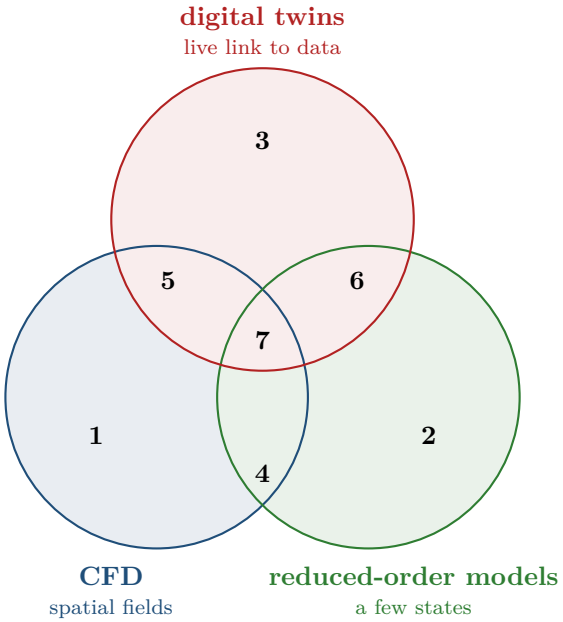
A reduced-order model simplifies the same physics to a small set of states. It keeps, for example, one temperature for each cold aisle, one for each rack exhaust and one for each hot aisle, with flows of air and heat between them, perhaps twenty states for a hall instead of millions of cells. Its output is a handful of numbers per time step, each representing a region rather than a point.

A digital twin is a different kind of object. It is a live model of the real facility, continuously updated with operational data. The term says nothing about how detailed the physics inside the model is; it says that the model is connected to a particular facility, receives its measurements and is reconciled with them. The three categories therefore overlap rather than exclude one another: a twin can be built on a reduced model or, at much greater cost, on CFD, and a reduced model can be derived from CFD.

22.1.2 Overlap and purpose

Figure 22.1 draws the three categories as overlapping sets. Each circle is one category, named outside it with the property that defines it: spatial fields for CFD, a few states for reduced-order models, a live link to data for digital twins. The seven numbered regions are the possible combinations, listed on the right. Region 1 is the classical CFD design study, run once for one geometry. Region 6, a twin built on a reduced network model, is where most operational twins

sit, for reasons the rest of the chapter develops. Region 7, a twin that runs a surrogate fitted to CFD, combines all three: the spatial knowledge of CFD, the speed of a reduced model and the link to the facility.



1 CFD design study, offline, one geometry
2 reduced model used offline for planning
3 twin built on data-driven or equipment models
4 surrogate or reduced spatial model fitted to CFD
5 CFD-based twin, updated from data at high cost
6 twin built on a reduced network model, the usual case
7 twin running a surrogate derived from CFD

Figure 22.1: The three kinds of model as overlapping categories. CFD and reduced-order models differ in spatial detail; a digital twin is defined by its connection to the operating facility, so it can contain either. The numbered regions are listed on the right.

The three kinds serve different purposes, and the difference can be stated along three axes. Along the axis of spatial detail CFD is strongest: it can say where in a room a hot plume forms. Along the axis of speed the reduced model is strongest: it runs in milliseconds and can be evaluated the hundreds of thousands of times that a probability or an optimization requires. Along the axis of connection to the real system the twin is strongest by definition: it is the only one of the three whose state follows the facility as it changes. No single model is strongest on all three axes, and the practical architecture that the chapter arrives at uses each where it is strong: high-fidelity models to understand the physics, reduced models for repeated inference, and live telemetry to keep the state current.

22.2 Computational fluid dynamics

This section states the equations a CFD model solves, describes the spatial detail they provide and where that detail is valuable, works the example of two racks that a lumped model cannot tell apart, and then sets out the costs: inputs, computing time, boundary conditions and a tendency to be underdetermined by the sensors of a real facility.

22.2.1 The conservation equations

A CFD model of a data hall solves three conservation laws at every point of the air volume. Write ρ for the air density (kg/m^3, about 1.2 at room conditions), $\mathbf{u}$ for the air velocity vector (m/s, with three components), p for the pressure (Pa), T for the temperature and t for time. The operator ∇ is the vector of spatial derivatives $(\partial/\partial x, \partial/\partial y, \partial/\partial z)$; $\nabla\cdot$ applied to a vector is its divergence, the net outflow per unit volume; and ∇^2 is the Laplacian, the sum of the three second derivatives, which measures how much a quantity at a point differs from its average

around the point. The pressure p here is a field in pascals, not a probability density; the two never appear together in this section.

Mass. Air is neither created nor destroyed:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\, \mathbf{u}) = 0. \tag{22.1}$$

In steady flow the first term vanishes and $\nabla \cdot (\rho \mathbf{u}) = 0$: as much mass leaves every small volume as enters it. Air in a data hall moves at a few meters per second, far below the speed of sound, and its density changes by only a few percent between the cold and the hot aisle, so it is usually treated as incompressible, ρ constant, which reduces the mass equation to $\nabla \cdot \mathbf{u} = 0$.

Momentum. Newton's second law for a small parcel of air reads

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\, \mathbf{u} \right) = -\nabla p + \mu\, \nabla^2 \mathbf{u} + \mathbf{f}. \tag{22.2}$$

The left side is mass per unit volume times acceleration: the change of velocity at a fixed point plus the change carried along by the flow itself, the term $(\mathbf{u} \cdot \nabla)\mathbf{u}$ that makes the equations nonlinear. On the right, $-\nabla p$ pushes air from high to low pressure; $\mu \nabla^2 \mathbf{u}$ is viscous friction, with μ the dynamic viscosity of air (Pa s, about 1.8×10^{-5}); and $\mathbf{f}$ is a body force per unit volume ($\mathrm{N/m^3}$), in a data hall chiefly buoyancy, the upward push on air warmer than its surroundings, which makes hot exhaust rise. Equations (22.1) and (22.2) together are the Navier–Stokes equations.

Energy. Heat is conserved as it is carried and spread:

$$\rho\, c_{p,\mathrm{air}} \left(\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) = k\, \nabla^2 T + q. \tag{22.3}$$

The left side is the rate of change of heat content of a parcel moving with the air, with $c_{p,\mathrm{air}}$ the specific heat of air. The term $\mathbf{u} \cdot \nabla T$ is advection, heat carried by the flow. On the right, $k \nabla^2 T$ is conduction, with k the thermal conductivity of air (W/(m K), about 0.026), and q is the volumetric heat source ($\mathrm{W/m^3}$), nonzero inside racks and equipment. An 80 kW rack modeled as a block of $0.6 \times 1.2 \times 2.0 = 1.44\,\mathrm{m^3}$ releases $q = 80\,000/1.44 \approx 5.6 \times 10^4\,\mathrm{W/m^3}$ inside that block and nothing elsewhere.

Turbulence. The ratio of inertial to viscous forces, the Reynolds number $\rho\, |\mathbf{u}|\, D/\mu$ for a flow of speed $|\mathbf{u}|$ past an object of size D, sets the character of the flow. For air at 2 m/s past a 0.5 m object it is $1.2 \times 2 \times 0.5/(1.8 \times 10^{-5}) \approx 6.7 \times 10^4$, far above the few thousand at which flows become turbulent. The eddies of a turbulent flow are smaller than any practical grid cell, so CFD codes add a turbulence model, an extra closure that represents the mixing those eddies cause. The turbulence model is an assumption, not a law, and it is one of the inputs on which the answer depends.

22.2.2 Spatial detail and its uses

The output of a CFD solution is three fields over the room: the temperature $T(x, y, z)$, the velocity $\mathbf{u}(x, y, z)$ and the pressure $p(x, y, z)$, where (x, y, z) are the coordinates of a point. This is what distinguishes it from every other model in the book. A lumped model states "the cold aisle is at 24°C". A CFD solution states "the top inlet of rack 18 reaches 31°C because a recirculation plume forms near the end of the aisle". The second statement names a location, a value and a mechanism, and it can be checked by placing a sensor at that location.

That kind of statement is what spatial problems need, and CFD is valuable wherever the question is spatial: the placement of racks in a row and of rows in a hall; the design of containment, the panels and doors that separate the cold aisle from the hot aisle; the placement and open area of perforated floor tiles; the analysis of recirculation, where hot exhaust finds its way back to inlets; the identification of hot spots; the cooling of high-density racks for artificial intelligence workloads, whose 50 to 100 kW concentrate the heat of several conventional racks in one footprint; the interaction of liquid and air cooling in racks that use both; and any physical change to the layout. In each case the answer depends on where air goes, and only a model that resolves space can give it.

22.2.3 Two racks in one row

Consider two racks of 80 kW each, illustrative, in the same cold aisle with the same supply temperature, $T_{\text{supply}} = 24°\text{C}$. With an air temperature rise of 15 K across each rack, the heat balance of Chapter 5, $\dot{Q}_{\text{rack}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, \Delta T$, gives each rack an airflow of

$$\dot{m}_{\text{air}} = \frac{80\,000\,\text{W}}{1000\,\text{J/(kg K)} \times 15\,\text{K}} = 5.33\,\text{kg/s}.$$

Rack A, in the middle of the row, receives only supply air, and its inlet is at 24°C from bottom to top. Rack B, at the end of the aisle, receives the same supply air plus hot exhaust from the neighboring row, which rises in the hot aisle, spills over the top of the row and curls into the upper part of the cold aisle. By the mixing rule of Chapter 6, $T_{\text{in}} = (1-\alpha)T_{\text{supply}} + \alpha T_{\text{hot}}$, a recirculation fraction α of exhaust at $T_{\text{hot}} = 24 + 15 = 39°\text{C}$ raises the inlet by 15α kelvin. A top inlet at 31°C corresponds to $\alpha = (31-24)/15 = 0.47$ at that height; a mid-height inlet at 26°C to $\alpha = 0.13$. The inlet of rack B is warmer than that of rack A, $T_{\text{in},B} > T_{\text{in},A}$, although the two racks have the same power, the same airflow and the same supply. The difference comes entirely from the geometry of the airflow.

Figure 22.2 draws the two racks in side elevation. In each panel the strip at the bottom is the underfloor plenum, the gap in the floor is the perforated tile, the gray box is the rack, the cold aisle is to its left and the hot aisle to its right. Blue arrows are supply air rising from the tile into the rack front; red arrows are exhaust leaving the rack back and rising. The numbers at the left are the inlet temperatures at three heights. In panel (a) all three are 24°C. In panel (b) the curved red arrow is the recirculation plume crossing over the row, and the inlet temperatures rise with height to 31°C at the top.

A lumped model with one node for the cold aisle cannot represent this difference. Its single cold-aisle temperature is an average over the aisle, and if the average is close to 24°C it predicts the same inlet for both racks; the warm top of rack B's inlet is invisible to it. The model is not wrong about the aisle; it is silent about the distribution of temperature within the aisle, which is exactly what the question about rack B turns on.

22.2.4 Inputs, cost and speed

A CFD model needs much more than equations. Table 22.1 lists what it requires. The first column names the input, the second says where it usually comes from, and the third says what happens to the answer if it is wrong. The list includes the geometry of the room, the locations and heat loads of the racks, the details of the raised floor and the ceiling, the locations and open areas of vents and tiles, the curves of every fan, the boundary conditions at every supply and return, and the choice of turbulence model.

The size of the computation follows from the grid. A hall of 20×15 m with a 4 m height, divided into cells of 10 cm, has $200 \times 150 \times 40 = 1.2 \times 10^6$ cells. Each cell carries three velocity components, a pressure, a temperature and typically two turbulence quantities, about seven unknowns, so the model solves for some 8×10^6 unknowns, and finer cells around racks and

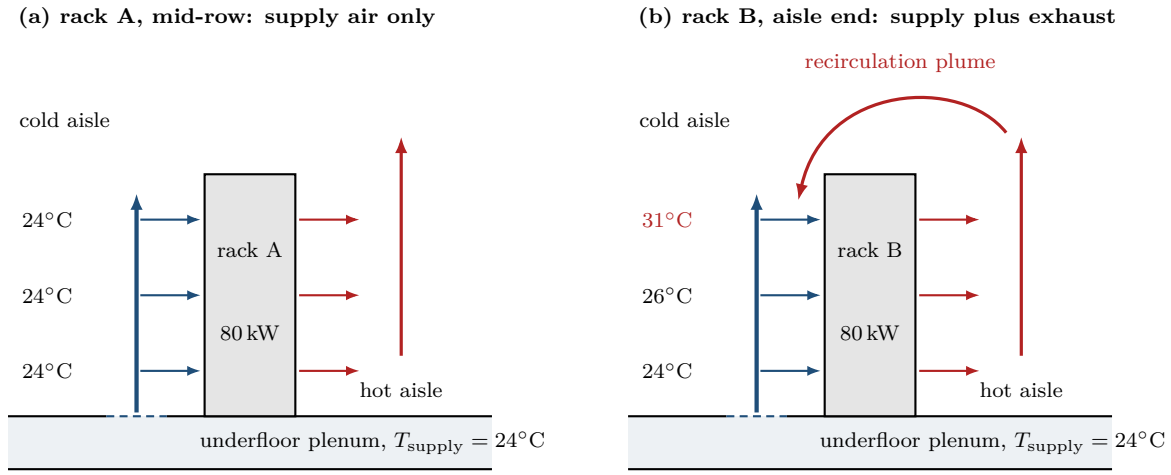


Figure 22.2: Two racks of the same power with the same supply air, illustrative. (a) Rack A in mid-row draws only supply air. (b) Rack B at the aisle end also draws exhaust that crosses over the row, and its inlet warms with height. A single cold-aisle node reports 24°C for both.

Table 22.1: Inputs of a CFD model of a data hall, their usual source, and the effect of an error in each.

Input	Usual source	Effect of an error
Room geometry, floor and ceiling	drawings, site survey	wrong flow paths; plenum pressure misplaced
Rack locations and blanking	asset database	plumes and recirculation in the wrong place
Rack heat loads P_{rack}	metered power, or nameplate	wrong source q; temperatures scaled wrongly
Tile and vent open areas	installation records	wrong distribution of supply air between racks
Fan curves and speeds	manufacturer data, BMS	wrong total and local flow
Supply temperature T_{supply}	CRAH sensors	offset of every temperature
Damper positions, leakage paths	rarely recorded	bypass and recirculation misjudged
Turbulence model	analyst's choice	mixing over- or under-predicted

tiles raise the number further. A reduced network model of the same hall has about 20 states. The ratio is of order 10^5 to 10^6, and it is paid at every evaluation.

The run time follows the size. A steady CFD solution of a hall takes minutes to hours on a workstation; this chapter uses ten minutes per run as an illustrative figure. That is acceptable for the question "should this aisle be redesigned?", which is asked once and answered with a handful of runs. It is unattractive for the question "what is the risk of exceeding the thermal limit, recomputed every 30 seconds?", which would need a run to finish twenty times faster than it does, and many runs rather than one. Spatial fidelity is bought with computational speed, and the two cannot both be maximized.

22.2.5 Boundary conditions and underdetermination

CFD solves the equations inside the room; everything that crosses the boundary of the room must be supplied. That includes the volumetric flow of every fan, the power P_{rack} of every rack, the supply temperature T_{supply}, the positions of dampers and the leakage paths through cable cutouts, gaps under racks and openings in containment. If these are wrong, the model produces a detailed answer that is wrong in detail. A field computed to a tenth of a kelvin on a centimeter grid, from a tile layout that was changed last year, is precise and incorrect. High model fidelity does not imply high truth fidelity: the resolution of a model says how finely it computes, not how closely its result matches the real room.

> **Principle 22.1.** *High model fidelity does not imply high truth fidelity. A detailed model computed from wrong boundary conditions gives a detailed wrong answer, and its resolution hides rather than reveals the error.*

The difficulty runs deeper than wrong inputs. Many of the inputs are not known at all: the resistance of every perforated tile, the area of every leakage path, the actual state of every fan. Suppose a hall has 200 tiles, 30 significant leakage paths and 12 fans, about 240 uncertain boundary parameters, and 40 rack-inlet temperature sensors. Many different combinations of the 240 parameters reproduce the 40 readings equally well, and each combination produces a different flow field. Fitting the CFD model to the readings is then a high-dimensional inference problem with far more unknowns than measurements, and the detailed model is underdetermined: the data do not choose among its possible fields. A more detailed model can therefore make identifiability worse, not better, in the sense of Chapter 15. Section 22.5.4 returns to this count.

22.3 Reduced-order network models

A reduced-order model replaces the fields of CFD by a network: a few nodes, each standing for a region of air or a piece of equipment, and edges between them that carry heat or air. This section writes the thermal network, the airflow network and their coupling, solves a small coupled example numerically, and states what the reduction gives up.

22.3.1 Nodes and edges

A hall reduced to about twenty thermal nodes follows the path of the air: cold aisle, rack, hot aisle, CRAH (computer room air handler) and back. Each node i carries a temperature T_i, the average over its region. Edges between nodes carry heat, by conduction through a partition or by the air that flows from one node to the next, or carry airflow driven by pressure differences. A hall with four cold-aisle zones, eight rack groups, four hot-aisle zones, a ceiling return, an underfloor plenum and two CRAHs has twenty nodes, which is the scale this chapter means by a reduced model of a hall.

The node is a control volume in the sense of Chapter 5: a region over which a balance is written. The choice of nodes is a modeling decision. A node that is too large averages away the very difference a question turns on, as the single cold-aisle node of Section 22.2.3 averaged away rack B. A node that is too small adds states that no sensor observes. A good reduced model puts node boundaries where measurements are and where the questions are: one node per zone that a decision distinguishes and a sensor can see.

22.3.2 The thermal network

Heat conducted between two nodes follows the thermal form of Ohm's law: the heat flow is the temperature difference divided by a thermal resistance,

$$\dot{Q}_{ij} = \frac{T_i - T_j}{R_{\mathrm{th},ij}}, \tag{22.4}$$

where $R_{\mathrm{th},ij}$ (K/W) is the thermal resistance of the edge between nodes i and j. Each node stores heat in a thermal capacitance C_i (J/K), and its balance is

$$C_i \frac{\mathrm{d}T_i}{\mathrm{d}t} = \sum_j \frac{T_j - T_i}{R_{\mathrm{th},ij}} + \dot{Q}_i, \tag{22.5}$$

where the sum runs over the neighbors of node i and $\dot{Q}_i$ is the heat generated at the node, for a rack node its electrical power. The equation says that the rate at which a node stores heat equals the heat flowing in from its neighbors plus the heat generated in it. It is exactly the equation of an electrical RC network, with temperature in the role of voltage, heat flow of current, thermal resistance of resistance and thermal capacitance of capacitance; a single node connected to a fixed temperature through one resistance relaxes with the time constant $\tau = R_{\mathrm{th}} C_{\mathrm{th}}$ of Chapter 8.

Collected over all nodes, equation (22.5) is a set of linear ordinary differential equations, one per node, which a computer integrates in microseconds per step. At steady state the left side is zero and the network reduces to a linear system: the temperatures are the solution of one matrix equation of the size of the number of nodes.

22.3.3 The airflow network

The airflow side has the same structure with different variables. Each node i carries a pressure p_i; each edge carries a mass flow $\dot{m}_{ij}$ from node i to node j; and the pressure drop along an edge grows with the square of its flow,

$$p_i - p_j = K_{ij}\, \dot{m}_{ij}\, |\dot{m}_{ij}|, \tag{22.6}$$

which is the system curve $\Delta p = K \dot{V}^2$ of Chapter 6 written for mass flow, with the sign of the flow kept so that a reversed pressure difference drives a reversed flow. Because $\dot{V} = \dot{m}/\rho$, the coefficient for mass flow is the coefficient for volumetric flow divided by ρ^2; here K_{ij} has units of Pa/(kg/s)^2. At each node, mass is conserved,

$$\sum_j \dot{m}_{ji} = \sum_k \dot{m}_{ik}, \tag{22.7}$$

the flow into the node from all upstream neighbors j equals the flow out of it to all downstream neighbors k. Fans are edges that raise pressure instead of dropping it, with the fan curve of Chapter 6 as their law. Equations (22.6) and (22.7) form a nonlinear network problem of the same kind as a direct-current resistor network, but with a quadratic law in place of Ohm's linear one; a pipe or gas network obeys the same equations.

22.3.4 The coupled network and a worked example

The two networks connect through the heat carried by moving air. When air flows from node i to node j, it carries its enthalpy with it, and the heat it delivers to node j, measured relative to node j's own temperature, is

$$\dot{Q}_{ij} = \dot{m}_{ij}\, c_p\, (T_i - T_j), \qquad \dot{m}_{ij} > 0. \tag{22.8}$$

Unlike conduction this transfer is one-directional: it enters the balance of the downstream node j, and the upstream node does not feel the temperature of the node it feeds. The chain of causes is

$$\text{pressure} \rightarrow \text{flow} \rightarrow \text{heat transport} \rightarrow \text{temperature},$$

and it is computed without solving the Navier–Stokes equations anywhere: pressures give flows by equation (22.6), flows give heat transport by equation (22.8), and heat transport gives temperatures by the node balances.

Figure 22.3 draws a small coupled network for the two racks of Section 22.2.3, with illustrative values. The boxes are equipment and the circles are air nodes. A CRAH fan, on the left, raises the underfloor plenum to a pressure p_1 above the room. Two tile paths, each standing for the perforated tiles in front of one rack, lead from the plenum to a cold-aisle zone in front of rack A and one in front of rack B, with resistances $K_A = 1.0$ and $K_B = 3.0\,\text{Pa}/(\text{kg/s})^2$; the tiles in front of rack B are partly blocked. Each rack's own fans draw a fixed $\dot{m}_{\text{rack}} = 5.33\,\text{kg/s}$ from its zone and push it into the hot aisle, which is at the room reference pressure $p = 0$ and returns to the CRAH along the bottom of the figure. The dashed edges close the mass balance of each zone: if the tiles deliver more than the rack draws, the surplus bypasses to the hot aisle (top); if less, the deficit is drawn back from the hot aisle as recirculation (bottom, red).

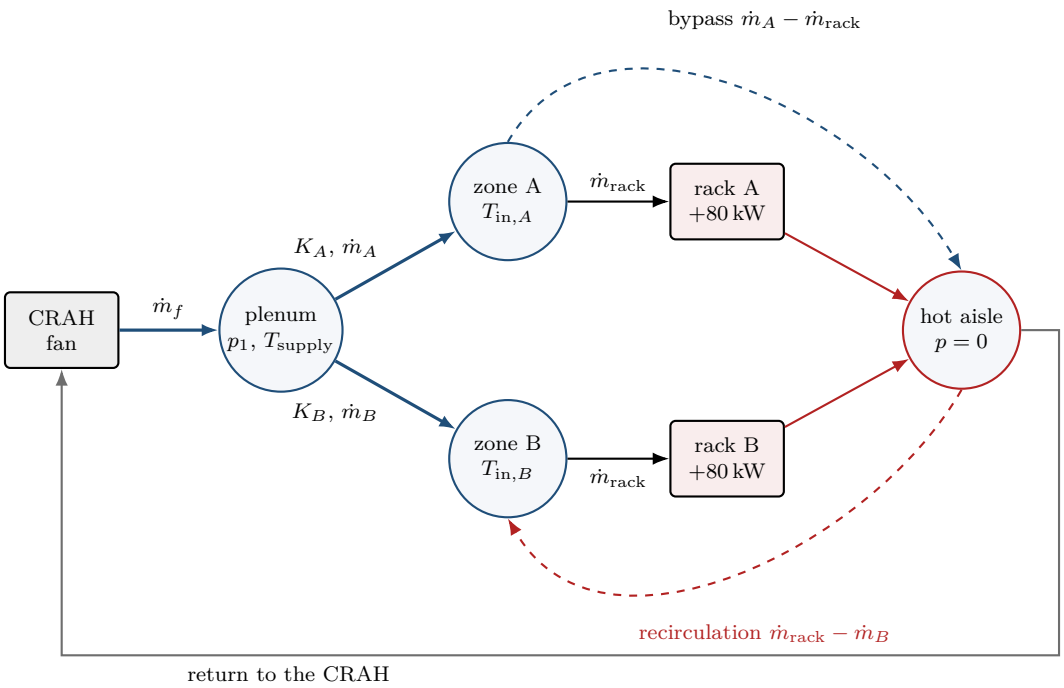


Figure 22.3: A coupled airflow and heat network for two racks, illustrative. Fan curve $p_1 = p_0 - K_f\,\dot{m}_f^2$ with $p_0 = 120\,\text{Pa}$ and $K_f = 0.5\,\text{Pa}/(\text{kg/s})^2$; tile paths $p_1 = K\,\dot{m}^2$ with $K_A = 1.0$ and $K_B = 3.0\,\text{Pa}/(\text{kg/s})^2$; each rack draws 5.33 kg/s for an 80 kW load and a 15 K rise. Dashed edges carry bypass (surplus) or recirculation (deficit).

Airflow. The only unknown pressure is p_1. The fan obeys the curve $p_1 = p_0 - K_f\,\dot m_f^2$, where $p_0 = 120\,\mathrm{Pa}$ is its shut-off pressure (the pressure it builds at zero flow) and $K_f = 0.5\,\mathrm{Pa/(kg/s)^2}$ the coefficient of its curve, so $\dot m_f = \sqrt{(p_0 - p_1)/K_f}$; each tile path obeys equation (22.6) with the room at zero, $\dot m_A = \sqrt{p_1/K_A}$ and $\dot m_B = \sqrt{p_1/K_B}$; and mass conservation at the plenum, equation (22.7), requires $\dot m_f = \dot m_A + \dot m_B$. The plenum pressure is therefore the root of the mass residual

$$g(p_1) = \sqrt{\frac{p_0 - p_1}{K_f}} - \sqrt{\frac{p_1}{K_A}} - \sqrt{\frac{p_1}{K_B}} = 0, \tag{22.9}$$

the flow the fan delivers minus the flow the tiles take. Newton's method solves it by the update $p_1 \leftarrow p_1 - g(p_1)/g'(p_1)$, with derivative $g'(p_1) = -\tfrac{1}{2}\big[(K_f(p_0 - p_1))^{-1/2} + (K_A p_1)^{-1/2} + (K_B p_1)^{-1/2}\big]$, the same iteration that solves a power flow or a pipe network. Table 22.2 lists the iterates from a starting guess of 40 Pa: the residual falls from 2.67 kg/s to 0.07, to 2×10^{-5} and to zero in four digits after three steps, at $p_1 = 53.48\,\mathrm{Pa}$. Because every law in this network is a square, the root also has a closed form for checking, $p_1 = p_0/(1 + K_f\,(K_A^{-1/2} + K_B^{-1/2})^2) = 120/(1 + 0.5 \times 1.5774^2) = 120/2.2440 = 53.48\,\mathrm{Pa}$; a network with more nodes has no closed form and is always solved this way.

Table 22.2: Newton iterations on the plenum mass residual of equation (22.9) for the network of Figure 22.3.

Iteration n	p_1 (Pa)	$g(p_1)$ (kg/s)	$g'(p_1)$ (kg/(s Pa))
0	40.000	2.673	−0.2038
1	53.119	0.0694	−0.1947
2	53.475	2×10^{-5}	−0.1946
3	53.476	$< 10^{-9}$	−0.1946

With $p_1 = 53.48\,\mathrm{Pa}$ the flows are

$$\dot m_A = \sqrt{53.48/1.0} = 7.31\,\mathrm{kg/s}, \qquad \dot m_B = \sqrt{53.48/3.0} = 4.22\,\mathrm{kg/s},$$

and the fan delivers their sum, $\dot m_f = 7.31 + 4.22 = 11.53\,\mathrm{kg/s}$. That is 11.53 kg/s against a total rack demand of $2 \times 5.33 = 10.67\,\mathrm{kg/s}$: there is a surplus of 0.87 kg/s for the pair. The split is uneven, however. Zone A receives 7.31 kg/s, 1.98 more than rack A draws, and the surplus bypasses. Zone B receives 4.22 kg/s, 1.11 less than rack B draws, and the deficit is drawn from the hot aisle.

Heat. Zone A's inlet is pure supply air, $T_{\mathrm{in},A} = 24°\mathrm{C}$. Zone B mixes 4.22 kg/s of supply air with 1.11 kg/s of exhaust, a recirculation fraction $\alpha_B = 1.11/5.33 = 0.208$. If the recirculated air is rack B's own exhaust, $T_{\mathrm{out},B} = T_{\mathrm{in},B} + 15\,\mathrm{K}$, the mixing rule gives $T_{\mathrm{in},B} = (1 - \alpha_B)T_{\mathrm{supply}} + \alpha_B(T_{\mathrm{in},B} + \Delta T)$, which solves to

$$T_{\mathrm{in},B} = T_{\mathrm{supply}} + \frac{\alpha_B\,\Delta T}{1 - \alpha_B} = 24 + \frac{0.208 \times 15}{0.792} = 27.9°\mathrm{C}, \tag{22.10}$$

and an exhaust at 42.9°C. Equivalently, $T_{\mathrm{in},B} = T_{\mathrm{supply}} + \Delta T\,(\dot m_{\mathrm{rack}}/\dot m_B - 1)$ whenever the tiles deliver less than the rack draws. A lumped model with one cold-aisle node sees only the totals: 11.53 kg/s supplied, 10.67 kg/s drawn, no deficit, both inlets at 24°C. The network places a 3.9 K excess on rack B and none on rack A.

Figure 22.4 shows how that excess depends on the blockage. The horizontal axis is the resistance K_B of rack B's tile path; the vertical axis is rack B's inlet temperature. The solid red curve is the network with the fan intact; the dashed red curve is the network after one of the

CRAH's two parallel fans is lost, which quadruples the fan coefficient to $K_f = 2.0\,\mathrm{Pa/(kg/s)^2}$; the gray line is the lumped model with the fan intact. With the fan intact the inlet stays at 24°C until K_B reaches $1.63\,\mathrm{Pa/(kg/s)^2}$, the point at which the tiles deliver exactly 5.33 kg/s, and rises steadily beyond it, to 27.9°C at $K_B = 3$ (the marked point) and 32.4°C at $K_B = 5$. The lumped model stays at 24°C across the whole range, because the total supply never falls below the total demand. The network finds an airflow-limited zone that the totals conceal.

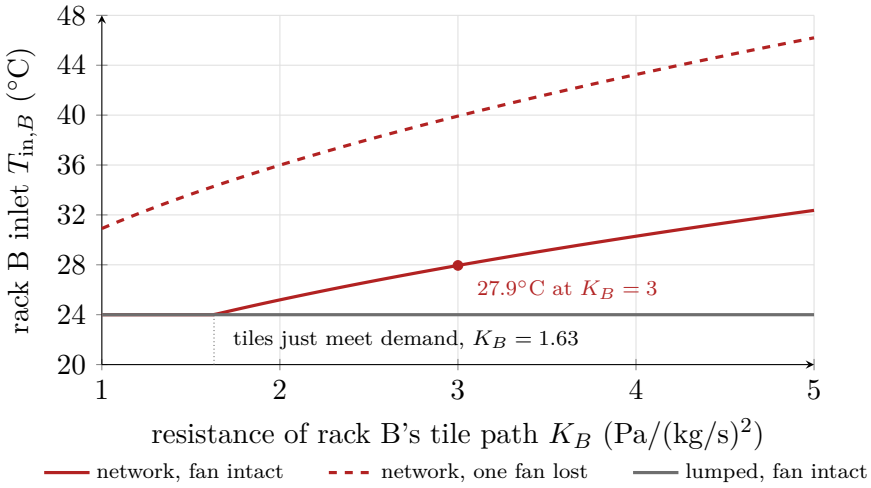


Figure 22.4: Inlet temperature of rack B against the resistance of its tile path in the network of Figure 22.3: $T_{\mathrm{in},B} = 24 + 15\,(\max(1, 5.33/\dot{m}_B) - 1)$ with $\dot{m}_B = \sqrt{p_1/K_B}$ and $p_1 = p_0/(1 + K_f(1 + K_B^{-1/2})^2)$, $p_0 = 120\,\mathrm{Pa}$, $K_f = 0.5$ (fan intact) or $2.0\,\mathrm{Pa/(kg/s)^2}$ (one of two fans lost). The lumped model compares totals only. Illustrative values.

22.3.5 What a reduced model gives up

The network buys its speed by giving up everything inside a node: local vortices, turbulent mixing, the shape and path of individual plumes, effects at the scale of a server or a cable bundle inside a rack, and precise spatial gradients. The recirculation edge of Figure 22.3 says how much exhaust returns to rack B's zone; it does not say where in the zone the exhaust goes, whether it enters the top inlet or is spread over the height. The network cannot say exactly where the hot plume forms. In the network of Section 22.3.4, 27.9°C is the mixed inlet temperature of rack B, an average over its face; the top inlet of the real rack may be several kelvin warmer, as in Figure 22.2.

What the network can say is which zone is airflow-limited and how much a zone's thermal margin changes when equipment changes. The loss of one fan in the example is such a question. With $K_f = 2.0\,\mathrm{Pa/(kg/s)^2}$ the plenum pressure falls to 20.1 Pa, the tile flows to 4.48 and 2.59 kg/s, and the inlets rise to 26.9°C for rack A and 39.9°C for rack B, while a lumped model would report 31.6°C for both. The same calculation at the scale of a hall answers "how much does hall B's thermal margin fall if CRAH 3 fails?": remove the CRAH's fan edge and cooling source, resolve the network, and compare. That is a question about zones and totals, and the network answers it in milliseconds.

22.3.6 Fidelity matched to three questions

Three questions about one hall show that the right model depends on the question. Whether moving one perforated tile will cool rack 18 depends on the jet of air from that tile and the plume near the rack, a spatial question for CFD. How hall B behaves after one CRAH fails

depends on the flows and heat balances of a few zones, a question for a reduced network. The probability that tomorrow's load and weather push any rack beyond its thermal limit depends on running the model across thousands of combinations of uncertain inputs, and a reduced model is preferable there because only a fast model can be run that many times.

The third question is the one that reverses the intuition that more detail is always better. A risk is a probability, estimated by repeated simulation over the uncertain inputs, in the manner of Chapter 17; Section 22.6.2 computes how many runs that takes and what they cost. A model too slow to be run that many times cannot produce the probability at all, however accurate each of its runs would be.

22.4 Digital twins and model updating

A digital twin is defined by how a model is connected to a real facility, not by the physics inside it. This section gives that definition and the architecture of a twin, separates the two kinds of updating a twin performs, explains why practical twins run reduced models or surrogates fitted to CFD, and describes a hierarchy in which detailed models are called only when the fast one fails.

22.4.1 A twin is defined by its link to data

A digital twin is a model of a specific physical asset linked to that asset's operational data. It may use CFD, reduced-order physics, machine learning, manufacturer equipment models, real-time sensor data, or several of these together. What makes it a twin is not the detail of its physics but three properties of its connection: it represents one particular facility, with that facility's equipment, topology and parameters; it receives that facility's measurements as they arrive; and it is reconciled with them, so that its state and parameters follow the facility as it changes. "Digital twin" is a statement about how the model is connected to the real system, not about how detailed the physics is.

A static simulation is not a twin. A CFD study run once at design time, not linked to telemetry and not reconciled with the plant as built and operated, is a simulation model, however detailed. It describes a facility as drawn on a given date. When the tile layout, the rack loads or the fan speeds change, it does not know, and its answers drift from the real room without any signal that they have done so. The same CFD model becomes part of a twin only when its inputs are refreshed from operational data and its outputs are compared with measurements.

> **Principle 22.2.** *A digital twin is defined by its connection to a real facility, not by the fidelity of its physics: a reduced network updated from telemetry is a twin, and a detailed CFD model never compared with measurements is not.*

22.4.2 Architecture of a twin

Figure 22.5 draws the architecture of a twin as a loop. The boxes are stages and the arrows carry information. On the left is the physical data center. Its sensors produce the measurement record $y_{1:t}$, all readings up to the present time t. The digital model receives the record and updates two things: its slowly varying parameters θ, by calibration, and its present state x_t, by state estimation. The updated model produces predictions, what-if scenarios and risks, which inform operator decisions. The decisions act on the facility as control actions u, and the loop repeats. Above the model, a dashed path brings offline CFD runs into the model as a surrogate, the subject of Section 22.4.4.

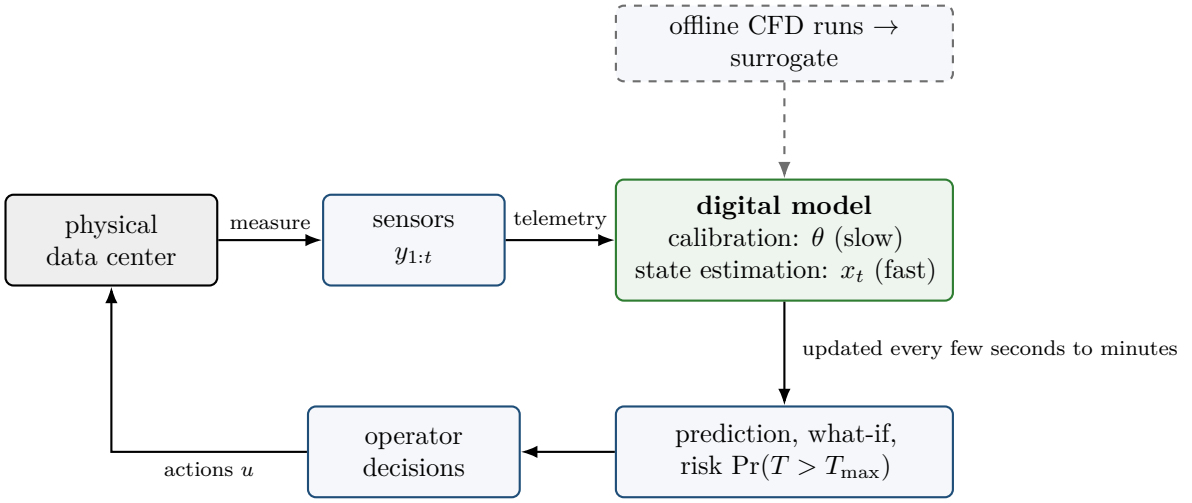


Figure 22.5: The architecture of a digital twin. Measurements update the model's parameters and state; the updated model predicts, evaluates scenarios and estimates risk for the operator, whose actions change the facility; offline CFD can enter as a surrogate (dashed).

The loop is the essential feature. A model that is run without the measurement arrow is a simulation; a model that receives measurements but whose outputs inform nothing is a monitoring display. A twin closes the loop repeatedly: physical facility, sensors, model, decisions, facility. Within the loop the three kinds of statement this book keeps apart remain distinct. The inlet temperatures and rack powers are measured; the rack heat, the air temperature rise and the cooling delivered are calculated from them by known relations; the recirculation fractions, the tile resistances K_{ij} and the coil conductances UA are inferred.

22.4.3 Calibration and state estimation

A twin updates two different kinds of unknown, and the distinction decides how each is handled. *Calibration* estimates parameters θ that are relatively fixed: thermal resistances R_{th}, coil conductances UA, airflow resistances K, fan curve coefficients. They change over weeks or months, as filters load or tiles are moved, and they are estimated from long stretches of data. *State estimation* tracks states $x(t)$ that change all the time: temperatures T, mass flows $\dot{m}$, the state of charge SOC of the UPS batteries, the heat stored in the room. They change over seconds to minutes and are estimated at every time step, as in Chapter 14. A twin needs both: correct parameters with a stale state predict the wrong present, and a current state propagated with wrong parameters predicts the wrong future.

Doing both together is model updating. At each time t the twin solves for the joint posterior of the present state and the parameters given all measurements so far, $p(x_t, \theta \mid y_{1:t})$, by the recursion

$$p(x_t, \theta \mid y_{1:t}) \;\propto\; p(y_t \mid x_t, \theta) \int p(x_t \mid x_{t-1}, \theta)\, p(x_{t-1}, \theta \mid y_{1:t-1})\, \mathrm{d}x_{t-1}. \tag{22.11}$$

The integral propagates the previous posterior through the model, the forecast step; the factor $p(y_t \mid x_t, \theta)$ weighs the forecast by how well it explains the new measurement, the update step. A model combined with data in this way is data assimilation, the method by which a weather service keeps its forecast model on the observed atmosphere. The Kalman filter of Chapter 14 is the special case of linear models and Gaussian errors.

The cost of updating CFD. Equation (22.11) must be evaluated every update interval, say every minute. For a reduced model with 20 states, a Kalman filter carries a 20×20 covariance

matrix of 400 numbers. For a CFD model with 8.4×10^6 unknowns, the corresponding covariance matrix would have $(8.4\times10^6)^2 \approx 7\times10^{13}$ entries, about 560 terabytes in double precision, before a single model run, and the forecast step needs at least one CFD run, ten minutes in the illustrative figure, per update. Updating a million-state model every minute under uncertain boundary conditions is therefore impractical, and practical twins run a reduced model, a surrogate or an emulator for the repeated inference, keeping CFD for occasional offline studies.

22.4.4 Surrogates derived from CFD

CFD and reduced models need not compete; the first can build the second. Run the detailed CFD model offline over many scenarios that span the expected operating range, varying the rack power P_{rack}, the CRAH fan speed N_{fan}, the supply temperature T_{supply} and the recirculation or containment configuration, and record the outputs of interest, such as the hottest rack inlet. Then fit a cheap function to the recorded pairs,

$$T_{\text{rack}} \approx \hat{f}(P_{\text{rack}},\ N_{\text{fan}},\ T_{\text{supply}},\ \alpha), \tag{22.12}$$

where the hat marks a fitted approximation of the simulator's input-output map. The CFD runs are expensive but done once; the fitted function is evaluated in microseconds whenever it is needed.

Such a function is a surrogate. Common forms are a response surface (a low-order polynomial in the inputs), a Gaussian process (a statistical interpolator that also reports its own uncertainty between and beyond the training runs), a neural network, and a reduced-basis model, which expresses the field as a combination of a few dominant spatial patterns extracted from the CFD runs and solves only for their weights. In every case the logic is the same, expensive simulator $\rightarrow$ cheap approximation, and the approximation is used where many evaluations are needed: in optimization, in Monte Carlo estimation of risk, and in real-time inference.

22.4.5 Domain of validity

A surrogate is reliable only where it was trained. A function fitted to CFD runs with $20°\text{C} < T_{\text{supply}} < 26°\text{C}$ has never seen what happens at 32°C; evaluating it there is extrapolation, and its answer there rests on the form of the fitted function, not on physics. The failure is not visible in the output: the surrogate returns a number at 32°C as readily as at 24°C. Section 22.5.6 shows, with numbers, how quickly the error of a fitted correction grows outside its training range.

Every reduced model has a domain of validity, not only surrogates. A network that represents rack recirculation as a fixed fraction is valid while server fan speeds stay in the range where that fraction was identified; a chiller curve is valid over the range of condenser temperatures in its test data. A twin should therefore carry the domain of validity of each of its models as explicitly as their parameters, and should flag, rather than silently answer, a query outside it. The operating conditions that matter most for risk, the hottest day, a failed unit, an unusual load, are exactly those most likely to fall outside the range a model was trained on.

22.4.6 Hierarchical modeling and escalation

The models of different fidelity can be organized as levels that are called in turn. Level 1 is a fast reduced model used for routine operation, run continuously. Level 2 is detailed CFD, called when the reduced model's residuals show an anomaly or its uncertainty is large. Level 3 is physical inspection of the facility, called when neither model explains the evidence. Each level is slower and more expensive than the one before, and each is used only when the one before has failed.

Figure 22.6 draws the escalation as a flow chart. The rectangles are the levels, the diamonds are tests. The reduced model predicts each measured quantity and computes the residual $r = y - \hat{y}$. While $|r|$ stays below a threshold r_{thr}, the loop on the left returns to routine operation. When the threshold is crossed, CFD is run with the current boundary conditions. If CFD reproduces the measurement, the anomaly is a spatial effect the reduced model lacks, and the missing mechanism is added to the reduced model (the path below). If CFD also disagrees, the cause lies outside both models, in the topology, the geometry or the sensors, and only a visit to the facility can find it.

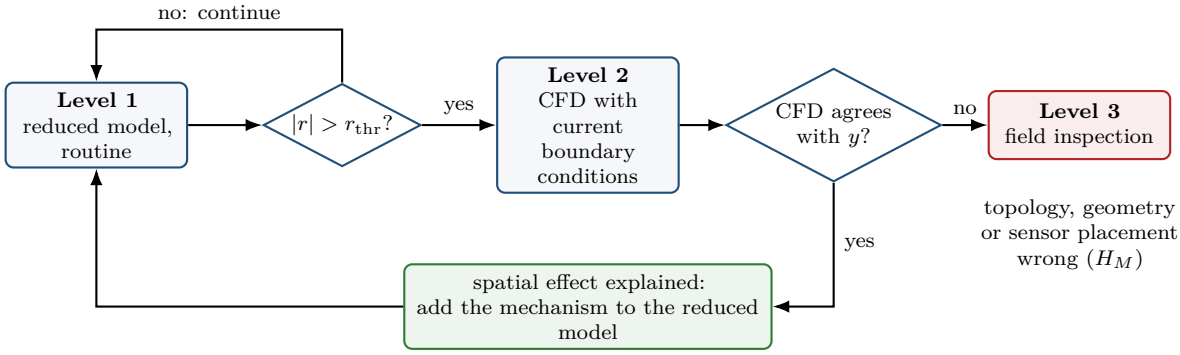


Figure 22.6: Hierarchical escalation between model levels. The fast model runs routinely; a large residual calls CFD; agreement of CFD with the measurement adds a mechanism to the fast model, disagreement sends the question to the facility itself.

A worked case shows the logic. The reduced model of a hall predicts the top inlet of rack 18 at 26.0°C; the sensor reads 31.0°C; the combined standard deviation of sensor and model is 0.5 K, and the threshold is set at three standard deviations, $r_{\text{thr}} = 1.5$ K. The residual of 5.0 K crosses it. CFD with the present rack powers, fan speeds and tile layout predicts 30.6°C at that point, within 0.4 K of the reading, and shows a plume over the end of the row: the reduced model lacked a recirculation edge into that zone, and one is added and its fraction inferred. Had CFD predicted 26.4°C instead, both models would disagree with the measurement by far more than their uncertainty, and the remaining explanations would lie outside the models: a missing blanking panel, a cable bundle blocking the tiles, a sensor mounted on the wrong rack. In the language of Chapter 16, the hypothesis H_M that the model itself is inadequate has become the most probable, and it is resolved by inspection, not by computation.

22.5 Model discrepancy, calibration and parameter count

Every model differs from the facility it represents, and the difference has to go somewhere. This section writes the difference as an explicit term, shows what happens to the parameters when it is left out, distinguishes calibrated from identified parameters, compares the number of parameters with the number of sensors, and describes how a reduced model is validated, corrected by a few high-fidelity runs, and combined with machine learning.

22.5.1 The discrepancy term

A reduced model predicts a rack inlet of 28°C; the sensor reads 30°C. Part of the 2 K may be measurement noise and part may be systematic, a mechanism the model does not contain. The two are separated by repetition. If the sensor's noise has a standard deviation of 0.3 K, a single 2 K gap is already 6.7 standard deviations; and if sixty readings over an hour average 30.0°C, the noise in the average is $0.3/\sqrt{60} = 0.04$ K, and essentially all of the gap is systematic.

The measurement equation should say so. Instead of $y = h(x, \theta) + \epsilon$, write

$$y = h(x, \theta) + \delta + \epsilon, \tag{22.13}$$

where $h(x, \theta)$ is the model's prediction of the measured quantity from the state and the parameters, ϵ is the measurement noise, random and different at every reading, and δ is the model discrepancy, systematic, varying smoothly with operating conditions, and present even when θ is at its best value. Writing δ explicitly gives the model's own error a place in the inference, with its own prior: typically zero mean, a standard deviation of the size that validation against measurements or CFD has shown, and a correlation over time and between neighboring sensors, because a missing mechanism affects them alike.

22.5.2 Parameters that absorb model error

If δ is left out, the discrepancy does not disappear; the fit pushes it into the parameters. Take a reduced model of the rack's inlet air path, $T = T_{\text{supply}} + R_{\text{th}} P_{\text{rack}}$, with $T_{\text{supply}} = 24°\text{C}$ and $P_{\text{rack}} = 80\,\text{kW}$. The physically correct resistance gives the predicted 28°C: $R_{\text{th}} = 4\,\text{K}/80\,000\,\text{W} = 5.0 \times 10^{-5}\,\text{K/W}$. The extra 2 K comes, in this illustrative case, from warm air leaking from a neighboring row at a rate that does not depend on the rack's load. Fitted without δ, the model can reach 30°C only by raising the resistance to $6/80\,000 = 7.5 \times 10^{-5}\,\text{K/W}$, fifty percent above its true value.

The fit at 80 kW is perfect, and the parameter is misleading in two ways. It mispredicts other conditions: at 40 kW the true inlet is $24 + 2 + 2 = 28°\text{C}$ while the inflated model gives $24 + 3 = 27°\text{C}$; at 120 kW the truth is $24 + 6 + 2 = 32°\text{C}$ and the model gives 33°C. And it points at the wrong cause: an inflated thermal resistance reads as a degraded rack airflow, a clogged filter or a failed server fan, which an engineer may be sent to find and will not find. A good fit with a misleading parameter is worse than a visible misfit, because the misfit at least announces that something is missing.

22.5.3 Calibrated is not identified

The same danger applies to calibrating a model against CFD or against measurements. Suppose a reduced model's airflow resistances are tuned until its temperatures match those of a CFD run. The match does not prove that the resistances were what was wrong. The tuning may be compensating for missing recirculation, for wrong rack powers, for a biased sensor, or for an error in the geometry, any of which moves the temperatures in the same direction as a change in resistance. The calibrated resistances then carry all of those errors and are labeled as one.

The distinction is between a calibrated parameter and a physically identified one. A parameter is calibrated when some value of it makes the model fit the data. It is identified when the data determine it uniquely and the model structure is right, so that its fitted value is the physical quantity of that name. Calibration is a statement about the fit; identification is a statement about the parameter. A calibrated value can be used for prediction inside the conditions it was calibrated on; only an identified value can be used to diagnose, to say that the tile resistance has doubled and the tile should be cleared. The tests that separate the two are those of Chapter 15: whether the parameter is determined by the data, and whether the model with a discrepancy term still assigns the misfit to it.

> **Principle 22.3.** *A calibrated parameter is not an identified one. A parameter tuned until a model matches data absorbs every error that moves the output the same way, and can be read as a physical quantity only when the data determine it and the discrepancy has its own term.*

22.5.4 Parameter count against sensor count

Higher fidelity raises the number of parameters, and the number of sensors does not follow. Consider a detailed model with 1000 uncertain parameters, tile resistances, leakage areas, local heat loads, turbulence coefficients, calibrated against 50 temperature sensors. Linearized around its operating point, the model's predictions change with the parameters through the sensitivity matrix $J = \partial h/\partial\theta$, with 50 rows and 1000 columns. Its rank is at most 50, so there are at least $1000 - 50 = 950$ independent directions in parameter space along which the parameters can move without changing any predicted reading to first order. Along those directions the Fisher information $\mathcal{I} = J^{\mathsf{T}}\Sigma^{-1}J$, where Σ is the covariance of the measurement noise, is zero, and the posterior is just the prior. The unknown dimension far exceeds the measurement dimension.

More data of the same kind helps less than it seems. Readings taken at other operating points add rows to J, but the sensitivities of the same fifty sensors at nearby operating points are strongly correlated, and the added rows add little rank. A reduced network with 12 parameters, by contrast, faces the same 50 sensors with a 50×12 sensitivity matrix that can have full rank 12: every parameter direction moves some reading. The simpler model can therefore give the more trustworthy inference, provided it contains the mechanisms the question needs and carries its discrepancy explicitly. Detail that the data cannot determine adds uncertainty without adding knowledge.

22.5.5 Validation of a reduced model

A reduced model is validated against three references: measured data from the facility, high-fidelity simulation, and physical constraints. Measured data are the final test but cover only the conditions that have occurred and the locations that have sensors. High-fidelity simulation covers conditions that have not occurred, a failed CRAH, a new rack layout, at the price of its own boundary-condition uncertainty. Physical constraints hold everywhere: energy must balance between the IT load and the heat removed, temperatures must increase along the air path through a rack, flows must be non-negative where fans force them.

Against CFD, validation computes the error of the reduced model over a set of representative scenarios,

$$e = T_{\mathrm{ROM}} - T_{\mathrm{CFD}}, \tag{22.14}$$

with T_{ROM} the reduced model's prediction and T_{CFD} that of the CFD run for the same scenario. The statistics of e say how far the reduced model can be trusted, and they can themselves enter the uncertainty analysis. In the illustrative case of Section 22.5.6, thirteen CFD runs over $20\,^{\circ}\mathrm{C} \le T_{\mathrm{supply}} \le 26\,^{\circ}\mathrm{C}$ give a mean error of $-1.83\,\mathrm{K}$ and a standard deviation of $0.70\,\mathrm{K}$: the reduced model reads low by about $1.8\,\mathrm{K}$ on average, with $0.7\,\mathrm{K}$ of scatter. Without further correction, a risk analysis would add $1.83\,\mathrm{K}$ to its predictions and widen their uncertainty by $0.7\,\mathrm{K}$; with the correction of the next subsection, the remaining scatter is much smaller.

22.5.6 Multi-fidelity correction

A multi-fidelity model uses a cheap low-fidelity model for most evaluations and a few expensive high-fidelity runs to correct it. The correction is learned as a function of the operating conditions x,

$$\delta(x) = T_{\mathrm{CFD}}(x) - T_{\mathrm{ROM}}(x), \qquad T(x) \approx T_{\mathrm{ROM}}(x) + \hat{\delta}(x), \tag{22.15}$$

where $\hat{\delta}$ is a fitted approximation of the discrepancy and the second expression is the operational model. The discrepancy here is the negative of the validation error of equation (22.14). The reduced model carries the physics and is evaluated as often as needed; the CFD runs are spent only on learning the correction, which is usually smoother and simpler than the temperature itself.

As an illustration, a reduced model predicts the hottest rack inlet of a hall as $T_{\mathrm{ROM}} = T_{\mathrm{supply}} + 4.0\,\mathrm{K}$. Thirteen CFD runs at supply temperatures from 20 to 26°C give discrepancies between 1.1 and 3.2 K, larger at warmer supply because the servers' fans speed up and draw more recirculated air. A straight line fitted to them, $\hat{\delta} = 1.83 + 0.343\,(T_{\mathrm{supply}} - 23)$ kelvin, leaves a root-mean-square residual of 0.18 K, a quarter of the 0.70 K scatter of the uncorrected error. At $T_{\mathrm{supply}} = 24$°C the corrected model gives $28.0 + 2.17 = 30.2$°C, consistent with the 30°C reading of Section 22.5.1.

Figure 22.7 plots the discrepancy against the supply temperature. The shaded band is the training range, 20 to 26°C. Filled dots are the training CFD runs; the solid line is the fitted correction inside the training range and the dashed line its continuation beyond it; the dotted curve is the illustrative true discrepancy that generated the runs, unknown in practice; open red circles are six check runs at 27 to 32°C, outside the training range. Inside the band the line follows the dots to within 0.36 K. Outside it the true discrepancy curves upward and the line does not: the error is 1.0 K at 28°C and 3.4 K at 32°C, where the corrected model gives $36.0 + 4.9 = 40.9$°C and the check run 44.3°C. The correction is reliable exactly where it was trained, as Section 22.4.5 stated.

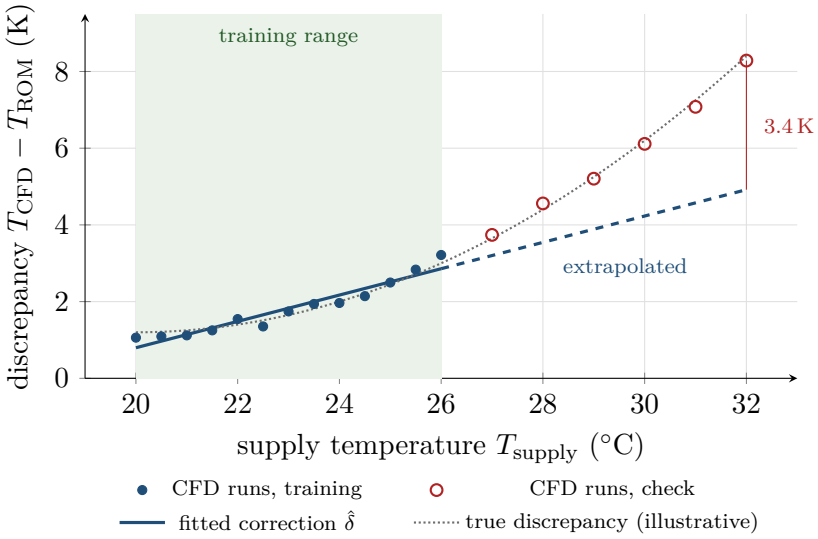


Figure 22.7: Multi-fidelity correction and its domain of validity, illustrative. Reduced model $T_{\mathrm{ROM}} = T_{\mathrm{supply}} + 4\,\mathrm{K}$; CFD discrepancy generated by the script `scripts/ch22_data.py` (seed 22) as $1.2 + 0.05(T_{\mathrm{supply}} - 20)^2$ plus noise of standard deviation 0.1 K; correction fitted to the training runs only, $\hat{\delta} = 1.83 + 0.343(T_{\mathrm{supply}} - 23)$. The error at 32°C is 3.4 K against 0.36 K at most inside the training range.

22.5.7 Physics-informed learning

Machine learning can enter a physical model in a more disciplined way than as a replacement for it. In physics-informed machine learning the conservation equations stay explicit, and a learned function supplies only the part the physics does not determine: a residual correction, such as $\hat{\delta}$ above; an unknown closure term, such as the recirculated heat; or a map of a parameter over operating conditions. For example, the recirculated heat flow into a zone can be written

$$\dot{Q}_{\mathrm{recirc}} = f_{\mathrm{ML}}(x),$$

a function learned from data of the zone's conditions x (rack powers, fan speeds, tile flows), and inserted into the node balance of equation (22.5), where energy conservation still holds exactly.

The difference from a black-box model is what is guaranteed. A neural network trained to map every input directly to every rack temperature can violate the energy balance, predict a negative recirculation or an inlet colder than the supply air, and extrapolate in any direction. The hybrid model can be wrong only in the learned term, whose range can be bounded ($0 \le \alpha < 1$ for a recirculation fraction), and the rest of its structure, the balances, the flow paths, the measured inputs, remains interpretable: an engineer can read off how much heat the model attributes to recirculation and check it against a CFD run or a temporary sensor.

22.6 Fidelity matched to the decision

The preceding sections described the models; this one chooses among them. It shows that the quantity to be inferred, not the physics, decides how much detail is needed; computes what uncertainty propagation costs at each fidelity; sets out the hierarchy of models and the level each operational question needs; places an inference layer above several models; and states the criterion for choosing.

22.6.1 The inference target decides

The question an operator asks determines which details matter. To estimate the probability $\Pr(T_{\text{rack}} > 35°\text{C})$ that a rack inlet exceeds a 35°C limit tomorrow, a model must represent the mechanisms that move that inlet temperature by a kelvin or more: the rack's power, its airflow, its recirculation, the supply temperature and their uncertainties. It does not need to represent every eddy in the aisle, because eddies change the inlet by fractions of a kelvin over seconds, and the probability is dominated by the uncertainties in load and weather, which are kelvins over hours.

Model complexity should therefore be driven by decision relevance: include a mechanism if leaving it out would change the decision, leave it out if it would not. The test is concrete. Run the decision with and without the mechanism, or with a cheap stand-in for it; if the recommended action and its risk do not change, the mechanism is not needed for this decision, however real it is. This is the same reasoning as the sensitivity analysis of Chapter 18, applied to the structure of the model instead of its inputs.

22.6.2 Cost of propagating uncertainty

A probability is estimated by Monte Carlo simulation: draw many samples of the uncertain inputs, run the model for each, and count the fraction of runs that violate the limit. Suppose 20 variables are uncertain, rack loads, supply temperatures, recirculation fractions, chiller capacities, and the risk to be estimated is of order 10^{-3}. The number of runs needed does not depend on the number of variables, but on the probability: the relative standard error of an estimated probability Pr from N_{MC} runs is $\sqrt{(1-\Pr)/(N_{\text{MC}}\Pr)}$, which for $\Pr = 10^{-3}$ and $N_{\text{MC}} = 10^5$ is $\sqrt{0.999/100} \approx 0.10$, a ten-percent error. A grid over the 20 variables would be far worse: even three values per variable gives $3^{20} \approx 3.5 \times 10^9$ combinations.

The cost of 10^5 runs depends entirely on the model. At ten minutes per CFD run,

$$10^5 \times 600\,\text{s} = 6 \times 10^7\,\text{s} = 16\,700\,\text{h} \approx 694\,\text{days} \approx 1.9\,\text{years},$$

and even a hundred runs in parallel would take a week; the analysis is impractical. At 0.01 s per reduced-model run,

$$10^5 \times 0.01\,\text{s} = 1000\,\text{s} \approx 17\,\text{min},$$

which is entirely feasible for a day-ahead risk. A risk refreshed every 30 s needs more: $30/0.01 = 3000$ runs per processor core per refresh, so about 33 cores in parallel to reach 10^5, or fewer samples and a wider error. For CFD, 30 s is a twentieth of a single run.

22.6.3 The model hierarchy

The models of this book form a hierarchy of five levels. Table 22.3 lists them. The first column is the level, the second describes the model, the third gives the typical number of states, and the last two give an illustrative time per evaluation and the time for the 10^5 evaluations of Section 22.6.2. Level 0 is rules: thresholds and alarms on measured values, with no physical model at all. Level 1 is lumped models, single-zone energy balances such as a hall or a facility treated as one node. Level 2 is network models with several thermal and airflow nodes, as in Section 22.3.4. Level 3 is reduced-order spatial models, which keep a coarse spatial field or a few dominant spatial patterns fitted to CFD. Level 4 is full CFD.

Table 22.3: The hierarchy of model fidelity with illustrative evaluation times and the time for 10^5 Monte Carlo evaluations on one processor core.

Level	Model	States	Time per evaluation	10^5 evaluations
0	Rules: thresholds, alarms	none	1 μs	0.1 s
1	Lumped: single-zone energy balances	1 to 5	0.1 ms	10 s
2	Network: several thermal and airflow nodes	10 to 100	10 ms	17 min
3	Reduced-order spatial model	10^2 to 10^4	1 s	28 h
4	Computational fluid dynamics	10^6 to 10^7	10 min	1.9 years

Each step up the hierarchy multiplies the cost of an evaluation by a factor of one hundred or more and adds spatial detail and unknowns. Each step down gains speed and loses detail. There is no universally best level: rules are fast and blind to causes, CFD is detailed and too slow for probability, and every level in between trades one for the other. The levels are also not exclusive in one facility; a twin may run Level 2 continuously, call Level 3 for a few zones and keep a Level 4 study for the next layout change.

22.6.4 Model by question

The level a question needs follows from what the answer depends on. Table 22.4 pairs seven operational questions with the levels that fit them. The first column is the question, the second the suitable level, the third the reason. Questions about totals, whether the cooling plant has enough capacity, need only lumped or network models. Questions about which zone or rack is the constraint need a network or a coarse spatial model. Questions about why a specific rack is hot, or where containment should change, turn on the geometry of the airflow and need a spatial model or CFD. Questions that need repeated evaluation, a real-time probabilistic risk or an optimization of controls every five minutes as in Chapter 20, need a reduced model whatever their spatial content, because no other level is fast enough.

The table is not a ranking of models but a matching. The same facility needs several levels, and the skill lies in calling each for the questions it fits and in knowing, for each answer, which level produced it and therefore what it cannot see.

22.6.5 An inference layer above several fidelities

An inference framework for a data center is not a CFD solver and not a replacement for one. It uses one or more physical models as engines, a reduced network model, a surrogate derived from CFD, manufacturer equipment curves, the outputs of an external simulation, and handles what none of them does by itself: uncertainty in their inputs and parameters, ranking of hypotheses

Table 22.4: Operational questions and the model fidelity that fits each.

Question	Suitable level	Reason
Is the cooling plant capacity sufficient?	lumped or network (1–2)	depends on totals and plant curves
Which hall is the bottleneck?	network (2)	depends on flows and balances per hall
Which rack may overheat?	network or reduced spatial (2–3)	needs per-zone airflow and recirculation
Why is rack 18 hot?	reduced spatial or CFD (3–4)	depends on a local plume or flow path
Where should containment be modified?	CFD (4)	a geometric design question
Real-time probabilistic risk	reduced (1–3)	needs 10^4 to 10^5 runs per update
Optimize controls every five minutes	reduced (1–2)	needs many runs inside each interval

about causes, risk as a probability, what-if scenarios, and the selection of the next measurement. The physical models answer "given these inputs, what happens"; the inference layer answers "given these measurements, what is probably true, how sure is that, and what should be done".

Figure 22.8 draws the arrangement. The bottom row holds the engines: CFD and its surrogates, the reduced thermal and airflow network, electrical and battery models of the power path, and equipment curves, together with the telemetry y. The wide box in the middle is the inference layer, which evaluates the joint posterior $p(x, \theta, H \mid y)$ over states, parameters and hypotheses, and the posterior over candidate model structures. The top row holds what the layer produces: risk, diagnosis, what-if analysis and the choice of the next measurement, each of which feeds the operator's decision. The inference layer coordinates several simulators rather than being one simulator; each engine is called at the fidelity the question needs.

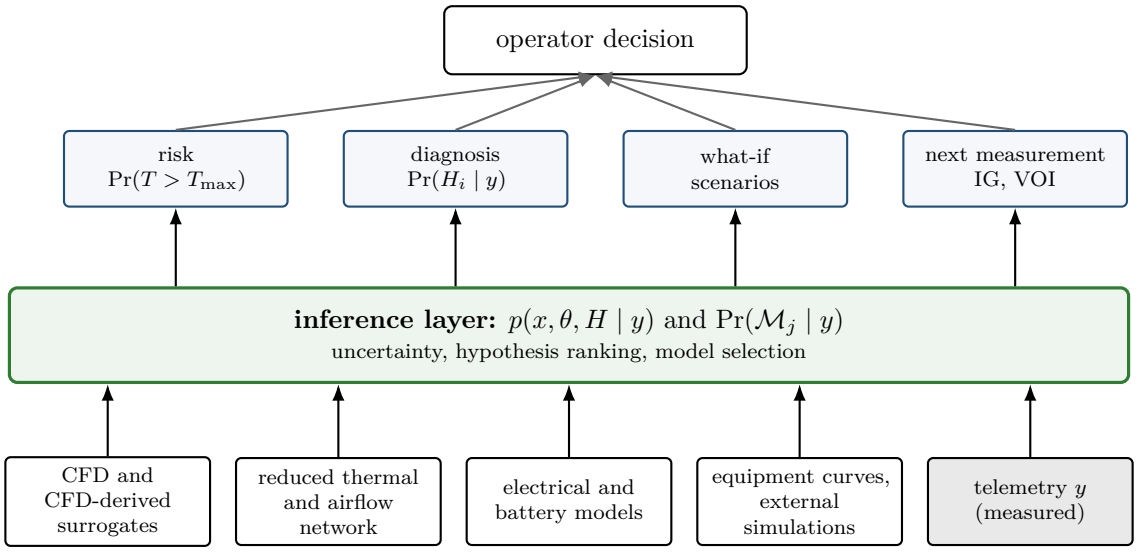


Figure 22.8: An inference layer above models of several fidelities. The engines in the bottom row compute physical consequences; the layer combines them with the telemetry into posteriors over states, parameters, hypotheses and model structures; the outputs in the top row feed the operator's decision.

Two consequences follow. First, the engines can be replaced or added without changing

the questions the layer answers: a better surrogate replaces an older one, a CFD study adds a recirculation mechanism to the network, and the risk, diagnosis and what-if outputs keep their meaning. Second, the uncertainty of each engine, its discrepancy δ and its domain of validity, enters the layer as information like any other, so that an answer that rests on an extrapolated surrogate is reported with the wider uncertainty it deserves.

22.6.6 Probabilistic model selection

The choice between fidelities can itself be made by inference. Let $\mathcal{M}_1$ be a simple airflow network without recirculation into a given zone and $\mathcal{M}_2$ an enhanced model with a recirculation term. Bayes' rule gives the posterior probability of each model given the measurements,

$$\Pr(\mathcal{M}_j \mid y) = \frac{p(y \mid \mathcal{M}_j)\,\Pr(\mathcal{M}_j)}{\sum_l p(y \mid \mathcal{M}_l)\,\Pr(\mathcal{M}_l)}, \tag{22.16}$$

where $p(y \mid \mathcal{M}_j)$ is the evidence for model j: the probability density of the observed data under the model, averaged over the prior of its parameters. A model with more parameters can fit more data sets, but it spreads its probability over all of them, so it assigns less to any one; the evidence therefore penalizes complexity that the data do not need. The rule is: keep the simple model if it is adequate, and escalate to the complex one when the evidence supports it.

A minimal illustration, with illustrative numbers, uses one measurement: the observed rise of a zone's inlet above the supply temperature, $y = T_{\text{in}} - T_{\text{supply}}$, with a measurement standard deviation of 0.5 K. Under $\mathcal{M}_1$ the rise is zero, so $y \sim \mathcal{N}(0, 0.5^2)$. Under $\mathcal{M}_2$ the rise is a recirculation effect with prior $\mathcal{N}(2, 2^2)$ kelvin, so $y \sim \mathcal{N}(2, 0.5^2 + 2^2)$, a standard deviation of 2.06 K. With equal prior probabilities, Table 22.5 gives the evidence and the posterior for two readings. The columns are the observed rise, the evidence under each model in units of 1/K, and the posterior of each model. On a normal day, $y = 0.3$ K, the simple model wins with probability 0.83, although the enhanced model, which contains the simple one as the case of zero recirculation, could also fit the reading. At the end of a row, $y = 4.8$ K, the simple model's evidence is 10^{-20} and the enhanced model is certain; its posterior for the recirculation rise is then 4.64 ± 0.49 K.

Table 22.5: Model selection between a network without recirculation ($\mathcal{M}_1$) and one with it ($\mathcal{M}_2$), equal priors, illustrative.

Case	y (K)	$p(y \mid \mathcal{M}_1)$	$p(y \mid \mathcal{M}_2)$	$\Pr(\mathcal{M}_1 \mid y)$	$\Pr(\mathcal{M}_2 \mid y)$
normal day	0.3	0.666	0.138	0.83	0.17
aisle-end rack	4.8	7.8×10^{-21}	0.077	≈ 0	≈ 1

22.6.7 The criterion

The criterion for choosing a model is not "how detailed is the model?" but "is the model sufficiently accurate for the decision?". The answer depends on five properties of the decision: its spatial scale (a facility total, a hall, a rack, a point on a rack face), its time horizon (seconds for control, hours for risk, years for design), the uncertainty already present in the inputs (a model more precise than its inputs adds nothing), the spatial detail the decision acts on (moving a tile, or shedding load in a hall), and the operational consequence of being wrong (an alarm acknowledged, or a hall shut down). A model is sufficient when its error, including its discrepancy and its uncertainty outside its domain of validity, would not change the decision.

Principle 22.4. *Choose the lowest model fidelity whose error would not change the decision, and state that error. Detail beyond that adds cost and unknowns without adding trust.*

This criterion explains the architecture that the chapter has built up. High-fidelity models are used to understand the physics and to train or correct reduced ones; reduced models are used for the repeated inference that risk, diagnosis and control require; live telemetry keeps their state and parameters current. The question to be asked of any facility model is not whether it is the most detailed available, but whether the right level of physics is used to make trustworthy inference and risk decisions under uncertainty.

22.7 What to retain

Definition 22.1 (Model fidelity for operation)**.** CFD provides spatial detail, a digital twin provides the live connection between a model and the real facility, and a reduced-order model provides speed and tractable inference; the operational architecture uses high-fidelity models for understanding, reduced models for repeated inference and live telemetry for state updating, and chooses for each decision the lowest fidelity that is accurate enough, with the model's discrepancy stated explicitly.

Table 22.6 lists the chapter's objects and the subsection that defines each.

Table 22.6: The objects of model fidelity and where each is defined.

Object	Expression	Role	Defined in
CFD equations	equations (22.1)–(22.3)	mass, momentum, energy in space	22.2.1
Boundary conditions	Table 22.1	inputs that set the answer	22.2.5
Thermal network	equation (22.5)	RC balance per node	22.3.2
Airflow network	equations (22.6), (22.7)	pressures and flows	22.3.3
Advective heat	$\dot{Q}_{ij} = \dot{m}_{ij} c_p (T_i - T_j)$	couples air and heat	22.3.4
Plenum residual	equation (22.9)	solved by Newton	22.3.4
Digital twin	Figure 22.5	model linked to data	22.4.1
Model updating	$p(x_t, \theta \mid y_{1:t})$	calibration plus estimation	22.4.3
Surrogate	$\hat{f}(P_{\text{rack}}, N_{\text{fan}}, T_{\text{supply}}, \alpha)$	cheap stand-in for CFD	22.4.4
Escalation	Figure 22.6	levels called in turn	22.4.6
Discrepancy	$y = h(x, \theta) + \delta + \epsilon$	model error given a place	22.5.1
Validation error	$e = T_{\text{ROM}} - T_{\text{CFD}}$	trust in a reduced model	22.5.5
Multi-fidelity model	$T_{\text{ROM}} + \hat{\delta}(x)$	cheap model, corrected	22.5.6
Model hierarchy	Table 22.3	levels 0 to 4	22.6.3
Model selection	equation (22.16)	escalate on evidence	22.6.6

22.8 Exercises

Exercise 22.1. A rack dissipates 80 kW with an air temperature rise of 15 K and a supply temperature of 24°C; take $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$. (a) Compute its airflow. (b) Its top inlet reads 31°C and its mean inlet 25.5°C. Compute the recirculation fraction at the top and on average, assuming the recirculated air is at the rack's exhaust temperature of 39°C. (c) In a CFD model the rack is a block of $0.6 \times 1.2 \times 2.0\,\text{m}$. Compute its volumetric heat source q. (d) Compute the Reynolds number of air at 1.5 m/s past a 0.6 m rack with $\rho = 1.2\,\text{kg/m}^3$ and $\mu = 1.8 \times 10^{-5}\,\text{Pa s}$, and say whether the CFD model needs a turbulence model.

Exercise 22.2. In the network of Figure 22.3, take $p_0 = 120\,\text{Pa}$, $K_f = 0.5$, $K_A = 1.0$ and $K_B = 4.0\,\text{Pa/(kg/s)}^2$. (a) Starting from $p_1 = 50\,\text{Pa}$, perform two Newton steps on

equation (22.9) and compare with the closed form. (b) Compute $\dot{m}_A$, $\dot{m}_B$, the bypass from zone A, the recirculation fraction of zone B and $T_{\mathrm{in},B}$. (c) What does a lumped model with one cold-aisle node predict? (d) With $K_B = 3.0$, compare two remedies: raising the fan's shut-off pressure to $p_0 = 150\,\mathrm{Pa}$, and clearing rack B's tiles to $K_B = 1.0$. Which removes the recirculation?

Exercise 22.3. A reduced model $T = T_{\mathrm{supply}} + R_{\mathrm{th}} P_{\mathrm{rack}}$ with $T_{\mathrm{supply}} = 22°\mathrm{C}$ is fitted at $P_{\mathrm{rack}} = 60\,\mathrm{kW}$ to a measured inlet of 28.0°C, of which 1.5 K is in fact a load-independent warm-air leak. (a) Compute the true R_{th} and the value fitted without a discrepancy term. (b) Compute the predictions of both at 30 kW and 90 kW and the truth at each. (c) A detailed model has 600 parameters and the facility 45 inlet sensors. What is the largest possible rank of the sensitivity matrix, and how many parameter directions leave every reading unchanged to first order? (d) Explain why the fit in (a) would pass a validation that used only data at 60 kW.

Exercise 22.4. (a) A day-ahead risk is estimated from 5×10^4 Monte Carlo runs. Compute the time on one core with a CFD model at 6 min per run, with 40 cores in parallel, and with a reduced model at 0.02 s per run. (b) If the risk is about 2×10^{-3}, what is the relative standard error of its estimate from 5×10^4 runs? (c) For the model selection of Table 22.5, compute $\Pr(\mathcal{M}_1 \mid y)$ for a reading $y = 1.5\,\mathrm{K}$ with equal priors. (d) Repeat (c) with prior $\Pr(\mathcal{M}_1) = 0.9$ and state whether the conclusion changes.

22.9 Connection to the TERRA framework

TERRA is an inference framework of the kind described in Section 22.6.5: it represents the facility as conservation laws on a graph and infers what the sensors do not read. The objects of this chapter map onto it as follows.

- **The reduced network is the graph.** The plenum, the cold-aisle zones, the racks and the hot aisle of Figure 22.3 are nodes; the tile paths, the rack airflows, the bypass and the recirculation are edges carrying mass flow under equation (22.6); mass conservation, equation (22.7), holds at each air node and the heat balance, equation (22.5) with the advective edges of equation (22.8), at each thermal node. The plenum residual solved by Newton's method in Table 22.2 is the kind of coupled balance the framework assembles and solves.

- **A twin is forward simulation plus posterior inference.** The loop of Figure 22.5 is the framework stepping the graph forward in time and updating $p(x_t, \theta \mid y_{1:t})$ from telemetry: temperatures and flows as states, tile resistances K_{ij}, recirculation fractions α and coil conductances UA as inferred parameters, each with a posterior rather than a calibrated point value.

- **Discrepancy and identifiability are checked, not absorbed.** The term δ of equation (22.13) and the hypothesis H_M of an inadequate model keep model error out of the parameters, and the rank count of Section 22.5.4 is the check the framework makes before it reports a parameter as identified rather than calibrated.

- **CFD and surrogates are components with bounds.** A surrogate fitted to CFD enters the graph as a component relation, and its training range, $20°\mathrm{C} < T_{\mathrm{supply}} < 26°\mathrm{C}$ in Figure 22.7, is a bound on that component: a query outside it is flagged and its uncertainty widened.

- **Model structure is a hypothesis.** A graph with and without the recirculation edge are the candidate models $\mathcal{M}_1$ and $\mathcal{M}_2$, ranked by equation (22.16) as in Table 22.5; the escalation of Figure 22.6 becomes adding an edge when the evidence supports it.

In summary, CFD resolves the air and temperature fields in space at a cost that rules it out for repeated evaluation, a reduced network model keeps a few nodes and edges and runs fast enough for probability and control, and a digital twin is any model kept connected to the facility by repeated updating of its state and parameters; the model for each decision is the lowest fidelity whose error, carried explicitly as a discrepancy and bounded by a domain of validity, would not change the decision, and an inference layer above several models chooses among them by the evidence. The next chapter turns to the data on which such models are built and tested, and to the public datasets that make data center modeling reproducible outside a single facility (Chapter 23).

Chapter 23

Data and Public Datasets

23.1 Public data and physical inference

Every estimate, diagnosis and risk statement in the preceding chapters rests on what is measured. A recirculation fraction is inferred from inlet, supply and exhaust temperatures; a coil conductance from air-side and water-side temperatures and flows; a capacity risk from a model whose parameters were fitted to telemetry. When the telemetry lacks a variable, the inference that needed it is not weaker; it is absent. The question of this chapter is

what data can actually be obtained for data centers, and what kinds of inference can those data support?

The answer has two parts. The first is a catalogue: which classes of public data exist, what a few representative datasets contain, and what each can and cannot support. The second is a method: how to judge a dataset by the physical balances its measurements allow to close, how each added measurement extends what can be inferred, how an inference method is validated when the true cause of an anomaly is rarely known, and how a study is designed from its question rather than from the data at hand. The descriptions of named datasets in this chapter follow a summary of their publishers' public documentation, not an independent inspection of the data, and should be checked against the current release before use.

Jargon: trace, telemetry, power capping, hyperscale, colocation, HPC, CFD, ground truth, label, aliasing, prediction interval, calibration

A *trace* is a time-stamped record of what a system did, such as the jobs submitted to a cluster and the resources each used; a *workload trace* records computation, a *power trace* records electrical power. *Telemetry* is the stream of measurements a facility collects from its equipment for operation. *Power capping* is the deliberate limiting of the power a server, rack or group of machines may draw. A *hyperscale* facility is a very large data center operated by one company for its own cloud services; a *colocation* facility leases space, power and cooling to many tenants. *HPC* is high-performance computing, the domain of supercomputers. *CFD* is computational fluid dynamics: the numerical solution of the equations of fluid flow and heat transfer on a fine spatial grid. *Ground truth* is the known true value or true cause against which an estimate is checked; a *label* is ground truth attached to a record, such as "this interval was a fan failure". *Aliasing* is the appearance of a fast variation as a false slow variation when a signal is sampled too coarsely. A *prediction interval* is a range that a method states will contain an unknown value with a given probability; the method is *calibrated* when its 95 % intervals do contain the value about 95 % of the time.

Key idea

A dataset is not judged by its size or its resolution but by which physical balances its measurements allow to close. Each measured variable adds a relation that can be tested or a quantity that can be inferred; each missing variable sets a boundary that no amount of the remaining data can cross. What a dataset can support is therefore an explicit output of an analysis, stated per question, before any inference result is reported.

23.1.1 Unevenness of public data

Public data on data centers exist, but they are uneven. There is a good deal of public data on workloads, on processor (CPU) and graphics processor (GPU) utilization, on power consumption and on temperatures. There is much less on the quantities the physical models of this book need to close their balances: detailed airflow, pressure differences, valve positions, chilled-water flow, equipment degradation, the full topology of the cooling system, failure events, and electrical, thermal and control telemetry recorded together on one clock.

The unevenness has a simple origin. The variables that are common in public data are the ones that computing systems record about themselves: a server knows its utilization and, increasingly, its own power and component temperatures, and a cluster scheduler knows its jobs. The variables that are rare belong to the facility: airflow, pressure and water flow are measured, if at all, by building systems that are separate from the computing systems, owned by a different team and rarely published. Equipment degradation and failure events are rarer still, because they are recorded in maintenance logs, if anywhere, and are commercially sensitive.

In the book's three-way distinction, public data are rich in *measured* compute quantities, from which the heat released by the IT equipment can be *calculated*, since nearly all IT electrical power becomes heat. They are poor in the facility quantities from which the state of the cooling system would be *inferred*: an airflow, a recirculation fraction or a coil conductance needs air-side and water-side measurements that most public datasets do not contain.

23.1.2 Four classes of dataset

Public data fall into four classes, ordered roughly from most to least common:

1. *Workload and compute* datasets: job traces, CPU and GPU utilization, memory use, job types and arrival times.
2. *Power* datasets: electrical power of servers, nodes, GPUs or groups of machines.
3. *Thermal and cooling* datasets: inlet and outlet air temperatures, component temperatures and, occasionally, air speed, ventilation data and room geometry.
4. *Integrated physical-system* datasets: electrical, air-side, water-side and control telemetry recorded together, with the topology that says which equipment serves which load.

The fourth class is the most valuable for physical inference, because it is the only one in which the energy balances of several subsystems can be written and checked against each other. It is also the rarest.

Figure 23.1 arranges the four classes from left to right. Each box names a class and lists what datasets of that class typically contain. The arrow above the boxes runs from the common classes toward the rare and more valuable one. The wide box beneath lists the variables that are rarely published in any class. Reading the figure from left to right follows the energy chain of the book: computation, then electrical power, then heat in the air, then the whole system; the further along the chain a variable lies, the less often it appears in public data.

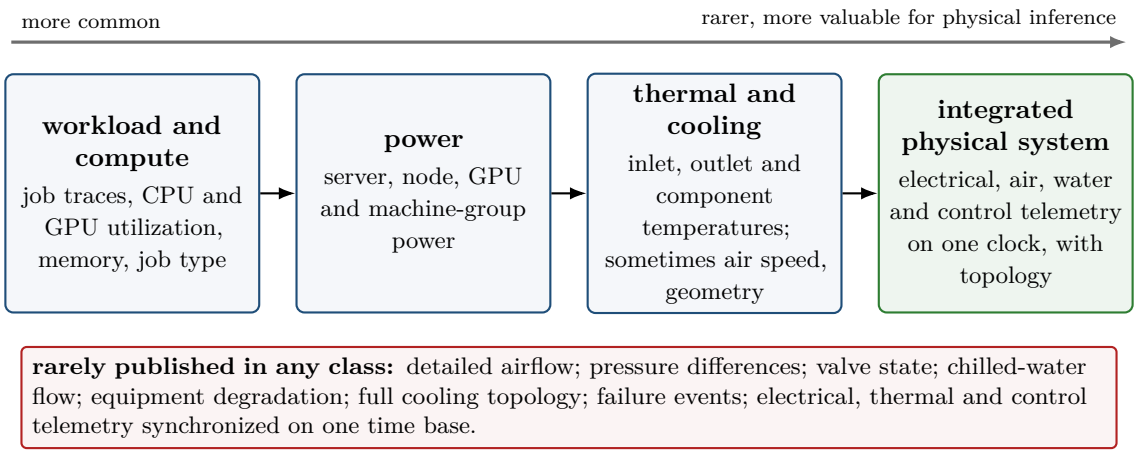


Figure 23.1: The four classes of public data center data, ordered along the energy chain from computation to the integrated physical system. The classes on the left are common; the integrated class on the right is the most valuable for physical inference and the rarest. The box below lists variables that are rarely published in any class.

23.2 Workload, power and thermal datasets

This section describes representative public datasets on one template: what each contains, what it supports, and what it cannot support. The descriptions are those of the publishers' documentation; the analysis of what each supports applies the physics of the earlier chapters.

23.2.1 Cluster workload and power traces

What it contains. The cluster-data repository published by Google is an example of the workload class. As described by its publishers, it contains Borg workload traces, and its 2019 release includes power traces for 57 power domains associated with Google data-center cells. It therefore combines the first two classes of Figure 23.1: records of computation and records of the electrical power drawn by groups of machines.

What it supports. With utilization $u_{\mathrm{CPU}}(t)$ and $u_{\mathrm{GPU}}(t)$ and power $P(t)$ on a common time base, a dataset of this kind supports three families of study. The first is forecasting: predicting $P(t+h)$ at a horizon h from the recent history of power and workload, which is the input to capacity planning and to the grid-interaction studies of Chapter 21. The second is the compute-to-power relation,

$$P = f\big(u_{\mathrm{CPU}},\ u_{\mathrm{GPU}},\ \text{memory},\ \text{job type}\big), \tag{23.1}$$

the empirical form of the linear server model of Chapter 2, $P \approx P_{\mathrm{idle}} + (P^{\max} - P_{\mathrm{idle}})\,u$, extended to more inputs. The third is aggregation and flexibility: how variable the workload is, how much the power of many machines averages out, and how much power could be shed or shifted by power capping. In short, such traces support workload variability, resource utilization, load forecasting, aggregation, compute-to-power relationships and power-capping studies.

What it cannot support. Workload and power traces are not a physical model of a cooling system. Without cooling measurements, the quantities that describe heat removal cannot be inferred. The rack heat balance of Chapter 5,

$$\dot{Q} = \dot{m}\, c_p\, (T_{\mathrm{out}} - T_{\mathrm{in}}), \tag{23.2}$$

has one known term, the heat $\dot{Q} \approx P$, and three unknowns; with no temperature and no flow, it constrains nothing. The airflow $\dot{m}_{\text{air}}$ and the coil conductance UA of Chapter 7 remain entirely undetermined, however long and fine the power record is.

23.2.2 AI training workload data

What it contains. A 2026 dataset published in the journal *Scientific Data* characterizes high-resolution AI training workloads on single and multiple GPU nodes. According to its documentation, it records power demand, CPU and GPU utilization, per-GPU power, memory usage and GPU temperature. Its authors note that ambient temperature and cooling configuration are not included.

What it supports, and what it does not. A dataset of this kind is excellent for compute-to-power behavior: how the power of a GPU follows the phases of a training job, how fast it changes, how correlated the GPUs of one node are, and how a multi-node job aggregates. These are the inputs to the electrical side of a facility model and to the load forecasts of the preceding subsection. It is weak for inference about facility cooling. A GPU temperature is a measurement at the end of a heat path that runs from the chip through its heat sink into the air and then into the room's cooling system; a reading at the end of that path, without the air temperature at its start or the airflow along it, cannot be attributed to any one part of the path.

23.2.3 An identifiability limit in GPU data

Suppose such a record shows GPU temperature T_{GPU} and GPU power P_{GPU} rising together. A natural reading is that the power rise caused the temperature rise. That may be so, but the air entering the server may also have warmed, or the cooling may have changed, for instance a fan speed or a supply temperature. Because neither the inlet air temperature nor the cooling state is recorded, the causal explanation is ambiguous.

The ambiguity can be made exact with the steady thermal-resistance model of Chapter 5,

$$T_{\text{GPU}} = T_{\text{in}} + R_{\text{th}}\, P_{\text{GPU}}, \tag{23.3}$$

where T_{in} is the temperature of the air entering the server and R_{th} the thermal resistance from the chip to that air. Take two illustrative observations: $P_{\text{GPU}} = 300\,\text{W}$ with $T_{\text{GPU}} = 60°\text{C}$, then $P_{\text{GPU}} = 600\,\text{W}$ with $T_{\text{GPU}} = 75°\text{C}$. Two explanations reproduce both observations exactly:

- *Power alone.* The inlet air stays at $T_{\text{in}} = 45°\text{C}$ and $R_{\text{th}} = 0.05\,\text{K/W}$: $45 + 0.05 \times 300 = 60°\text{C}$ and $45 + 0.05 \times 600 = 75°\text{C}$. All 15 K of the rise is due to power.
- *Power and inlet air.* $R_{\text{th}} = 1/30 \approx 0.033\,\text{K/W}$ and the inlet air warms from 50 to 55°C: $50 + 300/30 = 60°\text{C}$ and $55 + 600/30 = 75°\text{C}$. Only 10 K of the rise is due to power; 5 K is due to the air.

A third explanation, a fixed inlet temperature with a thermal resistance that rose because the cooling changed, can be constructed in the same way. The record contains two equations and, per observation, at least two unknowns; the share of the rise attributable to power is not identifiable from it. This is a hard inference boundary in a dataset that is high-resolution in every variable it does record: the resolution of the record does not substitute for the variable it lacks.

Figure 23.2 draws the causal structure. Each box is a variable; an arrow means "influences". Solid boxes are recorded in the dataset; dashed boxes are not. Utilization drives GPU power, and three arrows enter the GPU temperature: from GPU power, from the inlet air temperature and from the cooling configuration. Only one of the three parents of the measured temperature is itself measured, so the contributions of the other two cannot be separated from it.

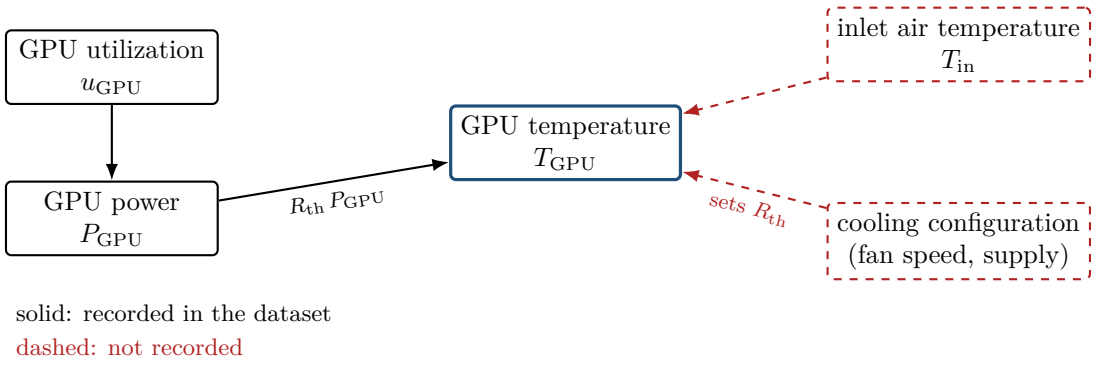


Figure 23.2: Causal structure of the GPU-temperature example, equation (23.3). The measured GPU temperature has three parents; only GPU power is recorded. A joint rise of power and temperature is consistent with several divisions of the rise among the three, so the cause is not identifiable from the record.

23.2.4 A supercomputer thermal dataset

What it contains. The thermal class is represented by the HazardNet dataset from the Marconi A2 Tier-0 supercomputer. As described by its publishers, it contains inlet temperatures, outlet temperatures and power consumption for 3312 computing nodes over roughly a year.

What it supports. Inlet temperature, outlet temperature and power together are the three measured terms of the heat balance, equation (23.2), for each node. A dataset of this kind therefore supports what workload and power traces cannot: the temperature rise across each node, its relation to power, a comparison of nodes against each other, and an approximate airflow, derived in the next subsection. Over roughly a year and thousands of nodes it also supports the study of slow changes and of differences between nodes that the balance can expose.

What it cannot support directly. On the contents listed, the airflow is calculated, not measured, and the record offers no independent airflow measurement against which that calculation could be checked. The airflow estimate is only as good as the assumptions behind it.

23.2.5 Airflow from inlet and outlet temperature

Equation (23.2) with the heat taken equal to the electrical power of the node, $\dot{Q} \approx P_{\mathrm{node}}$, solves for the mass flow,

$$\hat{\dot{m}} = \frac{P_{\mathrm{node}}}{c_p\,(T_{\mathrm{out}} - T_{\mathrm{in}})} = \frac{P_{\mathrm{node}}}{c_p\,\Delta T}. \tag{23.4}$$

For an illustrative node drawing 400 W with a 10 K rise and $c_p \approx 1.0\,\mathrm{kJ/(kg\,K)}$, $\hat{\dot{m}} = 0.4/(1.0 \times 10) = 0.04\,\mathrm{kg/s}$, about $0.033\,\mathrm{m^3/s}$ at an air density of $1.2\,\mathrm{kg/m^3}$. For an illustrative rack at 10 kW and a 12 K rise, $\hat{\dot{m}} = 10/(1.0 \times 12) = 0.833\,\mathrm{kg/s}$.

The estimate is calculated, and it rests on five assumptions:

1. nearly all the electrical power becomes heat carried away by the air, so that $\dot{Q} \approx P$;
2. the two temperature sensors represent the air entering and leaving along the flow path, not a local hot or cold pocket;
3. little air bypasses the equipment between the two sensors;
4. little exhaust air recirculates to the inlet sensor;

5. no unmodeled heat exchange takes place, such as conduction into the chassis or rack frame, or a liquid-cooled heat path.

Each violated assumption biases the estimate in a predictable direction. If 5 % of the power leaves by a path other than the air, the air carries $0.95P$ and $\hat{m}$ is high by $1/0.95 - 1 \approx 5.3\,\%$. If the outlet sensor sits where 20 % of the air it sees is bypass air at inlet temperature, it reads $0.8\,\Delta T$ and $\hat{m}$ is high by $1/0.8 - 1 = 25\,\%$. The estimate is also sensitive to temperature error, through the division by ΔT: at a 12 K rise, a 1 K error in the difference moves $\hat{m}$ from 0.833 to 0.769 or 0.909 kg/s, −7.7 or +9.1 %; at a 5 K rise, the same error moves it from 2.0 to 1.67 or 2.5 kg/s, −17 or +25 %. A small temperature rise makes a poor airflow estimate.

Figure 23.3 plots these effects for the 10 kW rack. The horizontal axis is the true temperature rise across the rack; the vertical axis is the airflow estimated by equation (23.4). The solid curve is the estimate with perfect sensors, which equals the true airflow. The shaded band is the range of estimates when the measured difference is off by up to 1 K in either direction. The dashed curve is the estimate when the outlet sensor sees 20 % bypass air. The band widens sharply toward small temperature rises, and the dashed curve lies 25 % above the true airflow everywhere: a sensor-placement error is a bias that no averaging removes. The estimate is useful, but not automatically trustworthy; it is a calculated quantity whose assumptions must be stated and, where possible, tested.

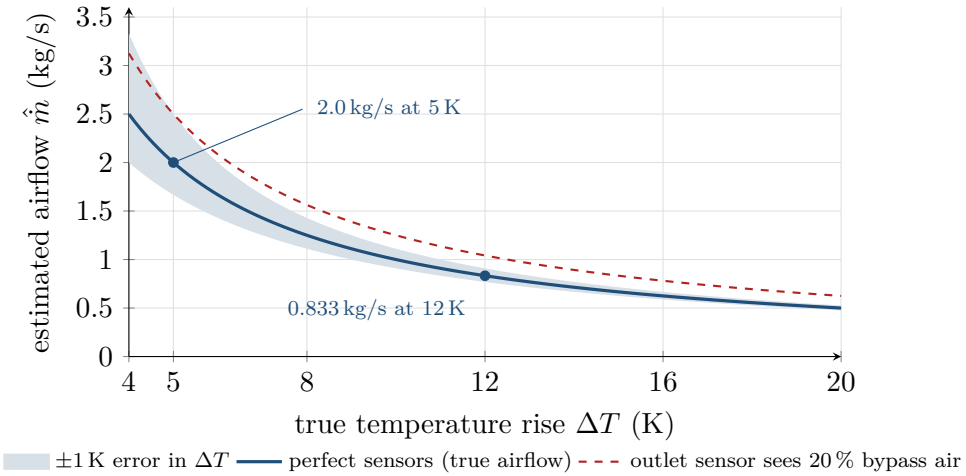


Figure 23.3: Airflow estimated from the heat balance, $\hat{m} = P/(c_p\,\Delta T)$ (equation (23.4)), for an illustrative rack with $P = 10\,\text{kW}$ and $c_p = 1.0\,\text{kJ/(kg K)}$. Band: $10/(\Delta T \mp 1)$; dashed: outlet reading $0.8\,\Delta T$, giving $12.5/\Delta T$. The estimate degrades sharply at small temperature rises and inherits any sensor-placement bias.

23.2.6 A server-room dataset with airflow and geometry

What it contains. A 2026 dataset from Aalborg University describes a server room. According to its documentation, it contains air speed, air temperature, server power, surrounding temperatures, ventilation data, the room geometry, the measurement locations and three-dimensional model files, with high-resolution measurements at roughly 1 to 10 s intervals and supplementary data at coarser resolutions. It was released with CFD and server-room measurement use in mind.

Why it is valuable. The dataset records several physical quantities of the same air path: power P, temperatures T and air speed v_{air}, together with where each was measured. That combination supports what neither of the preceding datasets can: a test of the heat balance

itself. The balance can be evaluated, as in the supercomputer case, to give an airflow from power and temperature; and the result can be compared with an airflow obtained independently from the measured air speed. The dataset contains both sides of the test.

23.2.7 Validation of the balance and of model complexity

The model-validation experiment that such a dataset allows has three steps. First, calculate the airflow from the thermal balance, $\hat{\dot{m}} = P/(c_p\,\Delta T)$, equation (23.4). Second, obtain the measured airflow from the air speed. A mean air speed $\bar{v}$ over a face of area A_{face} gives the mass flow $\dot{m}_{\text{meas}} = \rho\,\bar{v}\,A_{\text{face}}$, with air density ρ. Third, form the residual

$$r_{\dot{m}} = \dot{m}_{\text{meas}} - \hat{\dot{m}}. \tag{23.5}$$

With the illustrative 10 kW rack and a 12 K rise, $\hat{\dot{m}} = 0.833\,\text{kg/s}$. An illustrative mean air speed of 1.5 m/s over a 0.40 m^2 face at $\rho = 1.2\,\text{kg/m}^3$ gives $\dot{m}_{\text{meas}} = 1.2 \times 1.5 \times 0.40 = 0.72\,\text{kg/s}$, and the residual is $r_{\dot{m}} = 0.72 - 0.833 = -0.113\,\text{kg/s}$, about 14 % of the estimate. The residual does not say which side is wrong. It may be that part of the power leaves by another path, that the outlet sensor sees mixed air, or that a few air-speed readings do not represent the mean over the face. The value of the experiment is the question it makes answerable: under which conditions, at which loads, at which racks and in which periods, does the simplified thermal model fail?

Geometry. The spatial geometry and measurement points add a second question. With the room's shape and each sensor's location, a reduced model, such as the one-node rack balance or the zonal models of Chapter 22, can be compared against a CFD solution of the same room or against the measured spatial pattern of temperature and air speed. The question becomes one of model fidelity: how much spatial detail does a given operational question require? A rack-level energy check may need none; a recirculation diagnosis may need the position of every inlet sensor.

Table 23.1 collects the four datasets on the template of this section. Each row gives the class, the contents as documented by the publisher, what the contents support and what they do not. The last column is the chapter's analysis from the listed contents, not a statement by the publishers.

23.3 The ideal dataset and data quality

The catalogue shows what is available; this section describes what a dataset for physical inference would ideally contain, and three properties of data quality that matter as much as the list of variables: time alignment, sampling rate and missing variables.

23.3.1 The synchronized variable set

The ideal dataset for physical inference about a data center records, on one clock, seven groups of variables:

- *IT:* rack power P_{rack}, CPU and GPU utilization, and the workload;
- *air side:* T_{in}, T_{out}, the air mass flow $\dot{m}_{\text{air}}$ and the pressure difference Δp;
- *water side:* the chilled-water supply and return temperatures $T_{\text{CHW,s}}$ and $T_{\text{CHW,r}}$ and the water flow $\dot{m}_w$;
- *equipment state:* fan speed N_{fan}, valve position u_{valve} and chiller state;
- *electrical:* UPS power P_{UPS}, PDU power P_{PDU} and facility power P_{facility};

Table 23.1: Four representative public datasets on one template. Contents as described by the publishers' documentation; the support columns are the chapter's analysis.

Dataset (class)	Contents	Supports	Cannot support
Google cluster data (workload, power)	Borg workload traces; 2019 release with power traces for 57 power domains	workload variability, utilization, load forecasting, aggregation, compute-to-power models, power capping	a cooling-system physical model; $\dot{m}_{\text{air}}$, UA
AI training workloads, 2026 (workload, power, component temperature)	power demand, CPU and GPU utilization, per-GPU power, memory, GPU temperature; ambient temperature and cooling configuration not included	compute-to-power behavior of single and multi-node training	facility cooling inference; attribution of a GPU temperature rise
HazardNet, Marconi A2 (thermal)	inlet and outlet temperatures and power of 3312 nodes over roughly a year	temperature rise per node; airflow calculated under stated assumptions	validation of that airflow against a measurement
Aalborg University server room, 2026 (thermal, with geometry)	air speed, air temperature, server power, surrounding temperatures, ventilation, geometry, sensor locations, 3D models; about 1–10 s	heat-balance closure; measured against inferred airflow; reduced model against CFD	questions about the cooling plant or the electrical system, which need variables outside the listed contents

- *environment:* outdoor temperature T_{amb}, relative humidity RH and wet-bulb temperature T_{wb};
- *topology* G: which cooling and electrical components serve which racks.

With all seven groups, many more balances close than with any public dataset of the previous section: the rack heat balance $P_{\text{rack}} \approx \dot{m}_{\text{air}} c_{p,\text{air}} \, \Delta T$; the coil balance between air side and water side; the electrical balance $P_{\text{facility}} = P_{\text{IT}} + P_{\text{cool}} +$ losses of Chapter 3; the fan law between N_{fan}, Δp and flow of Chapter 6; and the chiller balance of Chapter 7. Each balance that closes is a test of the measurements in it, and each one that fails points to a sensor, a model or a fault.

Figure 23.4 draws the seven groups. The horizontal bar in the middle is the common time base. The boxes above and below it are the seven groups, each listing its variables, and the vertical ticks connect each group to the time base, meaning that every variable is time-stamped on the same clock. The topology box is shaded differently because it is not a time series but a structure, which itself changes when equipment is switched or maintained.

23.3.2 Time alignment

A dataset may contain every variable and still fail the balances if the variables are not synchronized. Suppose rack power is recorded every second, temperatures every five minutes and airflow every ten minutes. In steady operation the mismatch does not matter. In a transient, it does. Take an illustrative rack at 6 kW with $\dot{m}_{\text{air}} = 0.5$ kg/s and a 12 K rise, so that $0.5 \times 1.0 \times 12 = 6$ kW and the balance closes. At $t = 0$ the load steps to 10 kW. The power record shows 10 kW within a second. The last temperature reading still shows a 12 K rise and the last airflow reading 0.5 kg/s, so the calculated heat is still 6 kW and the balance residual is

$$r_Q = P - \dot{m}\, c_p\, \Delta T = 10 - 0.5 \times 1.0 \times 12 = 4\,\text{kW},$$

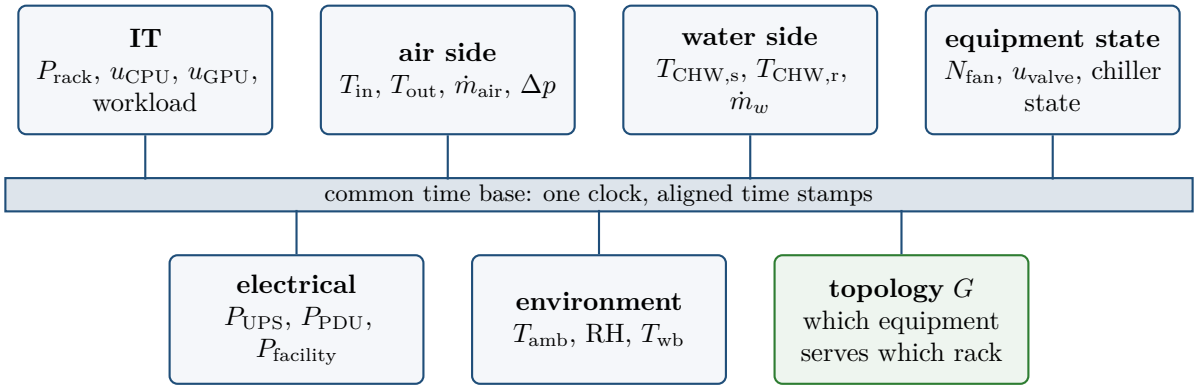


Figure 23.4: The ideal dataset for physical inference: seven groups of variables recorded on one time base. Topology (green) is a structure rather than a time series. With all seven, the rack, coil, electrical, fan and chiller balances can each be written and checked.

40 % of the power. The residual persists until the next temperature reading, up to five minutes later, and does not vanish even then, because the outlet temperature rises with the thermal time constant of Chapter 8 and because any change of server fan speed is not seen until the next airflow reading, up to ten minutes later. Read naively, the record says the heat balance failed by 40 %. Nothing failed: the variables were sampled at different instants.

Timing mismatch therefore looks like model failure, and data quality includes time alignment, not only sensor accuracy. Two remedies follow. The first is to evaluate the balance only over intervals long enough that all variables are averaged over the same window and the transient has passed. The second is to model the sampling explicitly: each reading is a measurement of the state at its own time stamp, and a dynamic model fills in the state between readings, as the state estimators of Chapter 14 do. Clock offsets between systems, which are common when the IT and facility data come from different software, must be measured and removed before either.

23.3.3 Sampling rate and aliasing

The sampling interval must match the dynamics to be inferred. Two illustrations show why.

A short event. A server fan event lasting 20 s, for instance a brief drop of fan speed, is invisible at a five-minute sampling interval. A five-minute record takes one reading every 300 s; an event of 20 s placed at a random time contains a reading with probability $20/300 \approx 6.7\,\%$, and even then only one reading, indistinguishable from a single bad value. Seeing such an event reliably needs several readings within it, an interval of a few seconds.

A periodic variation. Sampling theory states the general rule: a signal sampled every Δt_s seconds, a sampling frequency $f_s = 1/\Delta t_s$, can only represent variations slower than $f_s/2$, the Nyquist frequency. A faster variation is not merely lost; it reappears as a false slow variation at the aliased frequency

$$f_{\text{alias}} = \left| f - k\, f_s \right|, \tag{23.6}$$

where k is the integer that brings the result below $f_s/2$. Take an illustrative fan control loop that hunts, oscillating with a period of 310 s, recorded every 300 s. Then $f = 1/310$ Hz, $f_s = 1/300$ Hz, $k = 1$ and

$$f_{\text{alias}} = \frac{1}{300} - \frac{1}{310} = \frac{10}{93\,000}\ \text{Hz}, \qquad \text{an apparent period of } 9300\,\text{s} = 155\,\text{min}.$$

The record shows a smooth oscillation of the fan speed with a period of about two and a half hours, which an analyst could easily attribute to a slow load or weather cycle. The physical oscillation has a period of about five minutes.

Figure 23.5 shows both cases. The upper panel plots fan speed against time in minutes over three hours. The thin gray trace is the true speed, oscillating every 310 s so densely that it fills a band; the dots are the five-minute readings, and the dashed curve through them is the slow oscillation the record appears to show. The lower panel plots fan speed over fifteen minutes with a 20 s drop from 3000 to 1800 rev/min at about two minutes; the dots at the five-minute marks all read 3000 rev/min. In the upper panel the record shows a variation that does not exist at its apparent period; in the lower panel it misses one that does.

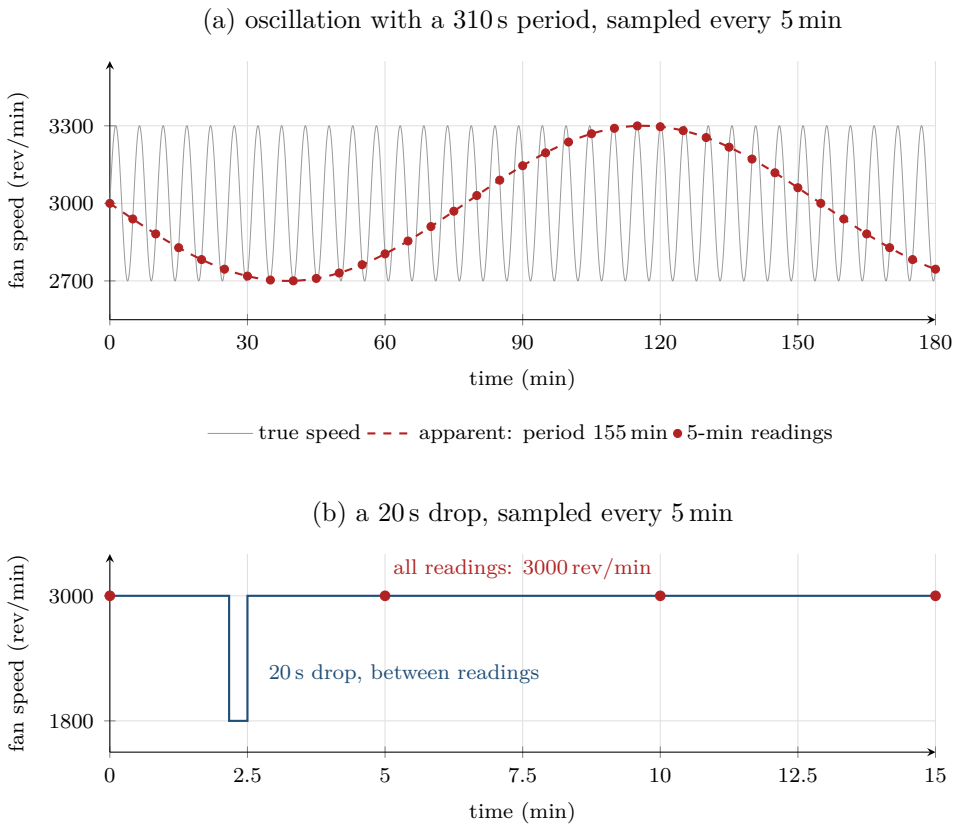


Figure 23.5: Sampling rate, illustrative. (a) True fan speed $3000 + 300\sin(2\pi t/310\,\mathrm{s})$ read every 300 s: the readings lie on $3000 - 300\sin(2\pi t/9300\,\mathrm{s})$, an aliased oscillation with a period of 155 min (equation (23.6)). (b) A 20 s fan-speed drop falls between five-minute readings and does not appear in the record.

The rule for thermal dynamics is the same. Thermal responses in a data center have time constants of minutes, as Chapter 8 showed; a response with an illustrative 4-minute time constant sampled every five minutes yields less than one reading per time constant, too few to fit the response, while a one-minute record yields four. A dataset is adequate for a question only if its sampling resolves the dynamics that the question is about.

Principle 23.1. *Time alignment and sampling rate are part of data quality. A balance evaluated on unsynchronized or undersampled variables fails for reasons that have nothing to do with the physics, and a fast variation sampled too slowly appears as a false slow one.*

23.3.4 Missing variables as inference boundaries

A missing variable does not add noise to an inference; it removes a question from the set that can be answered. With rack power, inlet temperature and outlet temperature, the airflow can be inferred, with the caveats of Section 23.2.5. A drop of that inferred airflow can be seen. Its cause cannot be named: a blocked filter and a degraded fan both reduce airflow, both raise the temperature rise at fixed power, and both leave P, T_{in} and T_{out} in the same relation. Separating them needs the pressure difference across the filter, Δp_{filter}, or the fan speed N_{fan}; without either, the two causes are not identifiable, in the sense of Chapter 15, from any length of record. The missing variables of a dataset are the boundaries of what can be inferred from it, and listing them is the first step of any analysis.

23.4 Capability as a function of instrumentation

What a dataset can support follows from which variables it measures. This section starts from the most common thermal configuration, rack power with inlet and outlet temperatures, grades what it supports, and then adds airflow, pressure, fan command and speed, and chilled-water telemetry one at a time, recording what each addition makes possible.

23.4.1 The capability map of power and air temperatures

Consider a dataset with P_{rack}, T_{in} and T_{out} and no direct airflow. Its capabilities, question by question, are:

- *rack heat load:* strong, because power is measured and nearly all of it becomes heat;
- *temperature rise:* direct, $T_{\text{out}} - T_{\text{in}}$;
- *approximate airflow:* possible, through equation (23.4) under its five assumptions;
- *abnormal thermal behavior:* possible, as a change in the relation between power and temperature rise;
- *diagnosis of low airflow:* weak, since the airflow is only calculated and a temperature-sensor error mimics a flow change;
- *blockage versus fan failure:* unsupported, as Section 23.3.4 showed;
- *recirculation:* usually weak, since recirculation raises T_{in} but without the supply and exhaust temperatures the fraction α cannot be computed;
- *thermal response prediction:* maybe, if the sampling resolves the dynamics and the parameters are identifiable;
- *cooling optimization:* limited, because neither the cooling equipment's state nor its power is recorded.

This list is the first column of Table 23.2. The following subsections fill the remaining columns.

23.4.2 Adding airflow

A measured airflow turns the airflow from a calculated quantity into a measured one and turns the heat balance into a test. The heat carried by the air, $\dot{Q}_{\text{meas}} = \dot{m}_{\text{air}}\, c_{p,\text{air}}\, \Delta T$, is now computed from three measurements and can be compared with the fourth, the power, through the energy residual

$$r_Q = P_{\text{rack}} - \dot{m}_{\text{air}}\, c_{p,\text{air}}\, (T_{\text{out}} - T_{\text{in}}). \tag{23.7}$$

When the model and the sensors are right, r_Q is close to zero. When it is not, its pattern points to a cause: recirculation, which raises T_{in} at the sensor without changing the heat; a measurement bias in any of the four sensors; an unmodeled heat transfer, such as heat leaving through the rack structure; or a timing mismatch of the kind in Section 23.3.2. The residual equation (23.7) is the airflow residual equation (23.5) multiplied by $c_{p,\text{air}}\,\Delta T$; the two express the same test in different units. With airflow measured, low-airflow diagnosis becomes strong and sensor bias becomes detectable, because one redundant relation now links four sensors.

23.4.3 Adding pressure

A pressure difference across a flow element, such as a filter, adds the flow-resistance relation of Chapter 6,

$$\Delta p \approx K\,\dot{V}^2, \tag{23.8}$$

with a resistance coefficient K that the data can now estimate as $K = \Delta p/\dot{V}^2$. A rising $K(t)$ means growing resistance to airflow, and blockage becomes identifiable. For an illustrative filter at $\Delta p = 60\,\text{Pa}$ and $\dot{V} = 2.0\,\text{m}^3/\text{s}$, $K = 60/2.0^2 = 15\,\text{Pa}\,\text{s}^2/\text{m}^6$. Months later, the same filter reads 75 Pa at $1.9\,\text{m}^3/\text{s}$, and $K = 75/1.9^2 = 20.8\,\text{Pa}\,\text{s}^2/\text{m}^6$, an increase of about 39 %. Flow fell and pressure difference rose, which is the signature of a resistance that grew; a weaker fan would have lowered both flow and pressure difference.

23.4.4 Adding fan command and speed

Two further signals, the commanded fan speed N_{cmd} and the actual speed N_{act}, separate the two causes of low airflow that Section 23.3.4 left joined:

- *blockage:* the fan behaves normally, $N_{\text{act}} \approx N_{\text{cmd}}$ at the expected command, while the resistance K rises and the flow falls;
- *fan degradation:* the command is normal, but the delivered speed falls short of it, or the speed is normal and the pressure or flow delivered at that speed is low, as for a worn impeller or a slipping belt.

Two inexpensive signals produce a large gain in diagnostic capability: blockage versus fan failure moves from unsupported to supported. Together with pressure, they also allow the fan curve of Chapter 6 to be checked: at a given speed, the fan laws predict the flow and pressure, and a persistent deficit at normal speed is a degradation of the fan itself.

23.4.5 Adding chilled-water telemetry

Chilled-water telemetry at a coil, the water flow $\dot{m}_w$ and the supply and return temperatures, gives the heat absorbed by the water,

$$\dot{Q}_w = \dot{m}_w\, c_{p,w}\,(T_{\text{CHW,r}} - T_{\text{CHW,s}}), \tag{23.9}$$

which can be compared with the heat given up by the air, $\dot{Q}_{\text{air}} = \dot{m}_{\text{air}} c_{p,\text{air}} (T_{\text{return}} - T_{\text{supply}})$ across the same coil, where T_{return} and T_{supply} are here the air temperatures entering and leaving the coil. In steady operation the two are equal. The comparison is a redundant energy-balance check: the same heat, computed from two independent sets of sensors.

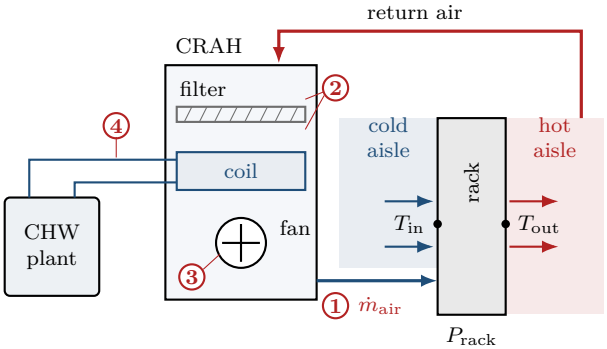
Table 23.2 lists the cumulative effect of the four additions. Its rows are questions; its columns are instrumentation configurations, each including everything to its left. The first column is the capability map of Section 23.4.1. The later columns follow from the relations of this section: airflow strengthens every airflow-related question; pressure begins to separate blockage; fan command and speed complete that separation; chilled-water telemetry adds a

second redundant balance and a view of the cooling side. Two questions, recirculation and thermal response prediction, improve only modestly along the row, because they need spatial temperature measurements or validated dynamics rather than more of the same air path.

Table 23.2: Capability as a function of instrumentation. Each column adds one group of measurements to all those on its left. First column: the capability map of Section 23.4.1; later columns: the chapter's analysis from equations (23.7)–(23.9).

Question	P_{rack}, T_{in}, T_{out}	+ airflow	+ pressure	+ fan command and speed	+ chilled water
Rack heat load	strong	strong	strong	strong	strong
Temperature rise	direct	direct	direct	direct	direct
Airflow	approximate	measured	measured	measured	measured
Abnormal thermal behavior	possible	residual test r_Q	residual test	residual test	two residual tests
Low-airflow diagnosis	weak	strong	strong	strong	strong
Blockage versus fan failure	unsupported	unsupported	partial (rising K)	supported	supported
Recirculation	usually weak	residual evidence	residual evidence	residual evidence	residual evidence
Sensor-bias detection	weak	possible	possible	possible	stronger
Air side against water side	unsupported	unsupported	unsupported	unsupported	supported
Thermal response prediction	maybe	better constrained	better constrained	better constrained	better constrained
Cooling optimization	limited	limited	limited	partial	partial, wider

Figure 23.6 places the added measurements on one air path. On the left is a computer room air handler (CRAH) drawn as a vertical box: return air enters at the top, passes through a filter, a chilled-water coil and a fan, and leaves at the bottom into the cold aisle; the rack sits between the cold and hot aisles, and the return path closes over the top. The coil's water pipes run to the chilled-water plant on the far left. The black labels at the rack are the baseline measurements; the circled numbers mark the four additions, and the column on the right states the relation each one adds.



baseline: P_{rack}, T_{in}, T_{out}; heat load and temperature rise; airflow only calculated, $\hat{\dot{m}} = P_{\text{rack}}/(c_{p,\text{air}}\Delta T)$

1 airflow $\dot{m}_{\text{air}}$: $r_Q = P_{\text{rack}} - \dot{m}_{\text{air}} c_{p,\text{air}} \Delta T$

2 filter Δp: resistance $K = \Delta p/\dot{V}^2$; rising K is blockage

3 fan N_{cmd}, N_{act}: fan fault against blockage

4 water $\dot{m}_w$, $T_{\text{CHW,s}}$, $T_{\text{CHW,r}}$: $\dot{Q}_w$ against $\dot{Q}_{\text{air}}$

Figure 23.6: The four added measurement groups of Table 23.2 placed on one air path, schematic. Each addition contributes one relation: an energy residual (1), a flow resistance (2), a command-versus-speed check (3) and a second, independent energy balance at the coil (4).

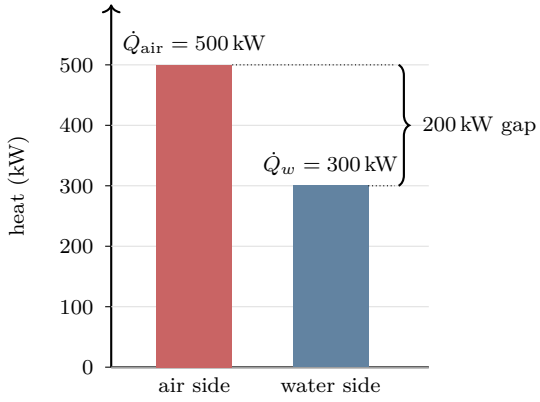
23.4.6 Redundant balances

A redundant balance earns its value when it does not close. Suppose the air side of a group of coils gives $\dot{Q}_{\text{air}} = 500\,\text{kW}$ and the water side gives $\dot{Q}_w = 300\,\text{kW}$. In steady operation the two should agree; they differ by 200 kW, 40 % of the air-side value. The difference has six candidate explanations:

1. a bias in the airflow sensor;
2. a bias in the water-flow sensor;
3. a bias in one or more temperature sensors;
4. transient heat storage, in the coil metal, the water in the pipes or the room air, during a change of load or setpoint;
5. sensors in the wrong locations, measuring air or water that is not the stream through the coil;
6. a missing heat path, such as a second coil or a heat source or sink between the measurement points.

The candidates differ in magnitude and in behavior over time, which makes the check informative. On the water side the temperature difference is small. With an illustrative $\dot{m}_w = 24\,\text{kg/s}$, the true heat corresponds to a water temperature difference of about 5 K, $24 \times 4.18 \times 5 \approx 500\,\text{kW}$; a reading of 3 K, a combined error of only 2 K in two water temperatures, gives $24 \times 4.18 \times 3 \approx 300\,\text{kW}$ and explains the entire gap. On the air side, an airflow reading high by a factor of $500/300 \approx 1.67$ would explain it as well, but a 67 % flow error is a much larger fault. A storage term decays as the transient passes, whereas a bias persists through every steady period; repeating the comparison over many steady periods therefore separates explanation 4 from the others.

Figure 23.7 draws the comparison. The two bars are the air-side and water-side heat, on a vertical scale in kilowatts; the bracket marks the 200 kW gap; and the list on the right gives the six explanations, each with the observation that would test it. The figure is a reminder that the most likely explanation of a large mismatch is not always the largest physical effect: a small error in a small temperature difference has a large effect on the computed heat.



candidate explanations and their tests

1. airflow-sensor bias: needs a 67 % flow error; check the fan law
2. water-flow-sensor bias: compare with the pump curve
3. temperature bias: 2 K in a 5 K water difference suffices; swap or recalibrate probes
4. transient storage: decays after the transient; repeat in steady periods
5. wrong sensor locations: inspect placement against drawings
6. missing heat path: walk the piping and air path for other exchangers

Figure 23.7: A redundant energy balance that does not close, illustrative: $\dot{Q}_{\text{air}} = 500\,\text{kW}$ against $\dot{Q}_w = 300\,\text{kW}$ at the same coils. Right: the six candidate sources of the gap and the observation that tests each. With $\dot{m}_w = 24\,\text{kg/s}$, a 2 K error in the water temperature difference accounts for the whole gap.

Principle 23.2. *Each measured variable is worth what it adds to the set of balances that can be closed. A measurement that completes a second, independent balance for the same heat turns a single estimate into a test.*

23.5 Ground truth and validation

An inference method is validated by comparing what it concludes with what was true. This section describes why that comparison is hard with public data, the three substitutes for labeled faults, how they combine into four stages of validation, and which claims a small dataset can nevertheless test.

23.5.1 Absence of labeled faults

Public datasets rarely say what an anomaly was. A temperature excursion in a record may have been an equipment failure, a sensor error, a workload change or a maintenance action, and the record does not say which. In the language of machine learning, the data are unlabeled. Supervised validation of fault diagnosis, in which a method's diagnosis of each event is scored against the known cause, is therefore difficult: there is nothing to score against. The few operators who keep maintenance logs rarely publish them, and the logs, when they exist, are written for repair, not for analysis, and record the action taken rather than the physical cause.

The consequence is that a diagnostic method can rarely be shown, from public data alone, to name real faults correctly. What can be shown is narrower: that the method detects departures from a validated normal model, and that it recovers known causes in the settings described in the next three subsections.

23.5.2 Synthetic faults

The first substitute is to inject faults into measured normal operation. A period judged normal, for instance one in which the balances close, is taken as a base, and a fault of known kind and size is added to it numerically. Two illustrative injections:

- a temperature-sensor bias $b_T = +3\,^{\circ}\mathrm{C}$ on the outlet temperature. With the 10 kW rack at a true 12 K rise, the recorded rise becomes 15 K and the calculated airflow $10/15 = 0.667\,\mathrm{kg/s}$ instead of $0.833\,\mathrm{kg/s}$, an apparent 20 % flow drop that did not happen;
- an airflow reduction $\dot{m}' = 0.7\,\dot{m}$. At constant power the rise becomes $12/0.7 = 17.1\,\mathrm{K}$, the outlet temperature rises by 5.1 K, and the calculated airflow falls to $10/17.1 = 0.583\,\mathrm{kg/s} = 0.7 \times 0.833\,\mathrm{kg/s}$.

The test is whether the method recovers what was injected: the right kind of fault, at the right sensor or edge, with the right size, and with a stated uncertainty that covers the injected value. The two examples also show why the test is demanding: with power and temperatures alone, a 3 °C outlet bias produces the same calculated signature as a 20 % flow drop.

Synthetic-fault validation must be kept distinct from real-fault validation. An injected fault has exactly the form the injector chose, usually the form the method assumes; a real fault may drift, interact with control loops, or affect several variables at once. Success on injected faults shows that the method works when its assumptions hold; it does not show that real faults satisfy them.

23.5.3 Controlled experiments and CFD-generated data

Controlled experiments. The second substitute is an experiment in a research server room. The experimenter changes a fan speed, modifies the airflow resistance, for instance by partly blocking a filter or a rack face, raises the server load, or shifts a cooling setpoint. Because the intervention is known, its time, kind and size are ground truth, and the physical response is real rather than injected. This is stronger validation than synthetic faults: the method is scored against real physics with a known cause. Its limits are scale and representativeness: a research room is small, and the interventions an experimenter can safely make are a subset of the faults that occur in operation.

CFD-generated data. The third substitute is a CFD model of a room, run to produce synthetic sensor data for conditions with known recirculation, blocked airflow or fan failure. A CFD model can generate large hypothesis libraries, many faults at many locations and sizes, which no experiment could produce. The risk is circularity: a method trained or tuned on CFD output may learn the CFD model's assumptions, its turbulence model, its boundary conditions and its idealized geometry, rather than the behavior of the real room. CFD-generated data must therefore be checked against real observations of the same room, which is exactly what a dataset with geometry and measured air speed, such as the server-room dataset of Section 23.2.6, makes possible.

23.5.4 Four stages of validation

The substitutes combine into four stages, each with its own setting and its own kind of evidence:

1. *development*, on synthetic physics models, where every state and every fault is known;
2. *basic validation*, on public datasets, where the measurements are real but faults are unlabeled, so the evidence is closure of balances and recovery of injected faults;
3. *controlled fault experiments*, in research facilities, where interventions are real and known;
4. *operational validation*, in commercial data centers, where the system is at full scale and complexity but ground truth comes only from operator records and later inspection.

Figure 23.8 draws the four stages from left to right. Each box names the stage and its setting. The two rows of notes beneath the boxes state, for each stage, what ground truth is available and how realistic the setting is. Realism grows from left to right. Ground truth does not shrink monotonically: it is complete in stage 1, largely absent in stage 2, known by design in stage 3, and partial in stage 4. No single stage both is realistic and knows the truth, which is why a method needs all four.

23.5.5 Representativeness and what small datasets can test

A dataset represents the system class it was recorded in. A university server room differs from a hyperscale cloud facility, a colocation facility, an HPC center and an AI training facility in rack density, cooling architecture, redundancy, workload patterns and controls. A result obtained on one class does not transfer to another without argument: an airflow estimate that closes within a few percent in a room with a single cooling unit and steady load says little about a facility with dozens of units under fast-changing AI workloads. Conclusions must be bounded to the system class of the data.

That limit does not make small datasets uninformative. Many of the claims a physical-inference method makes are methodological, and they can be tested wherever the physics holds. Two examples:

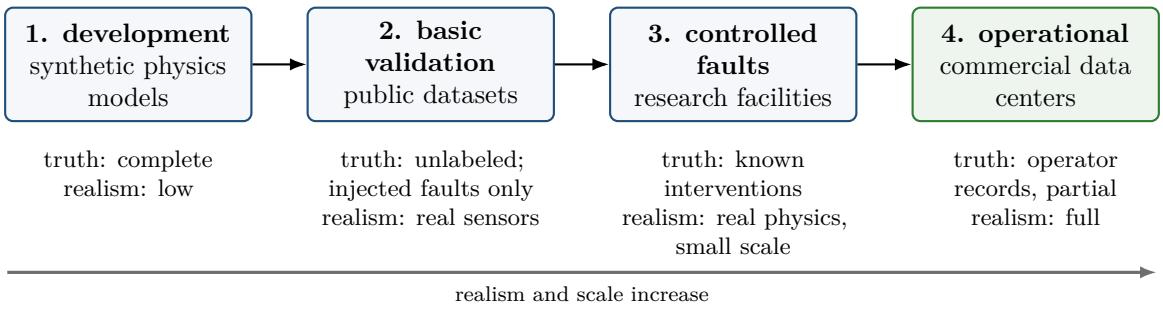


Figure 23.8: Four stages of validation for a physical-inference method. Realism increases from left to right; ground truth is complete only in the synthetic stage, absent in public data except for injected faults, known by design in controlled experiments, and partial in operation.

- whether it can be detected that a thermal balance does not close;
- whether a sensor bias can be separated from an airflow degradation when enough measurements exist.

Neither needs a hyperscale facility. The heat balance is the same equation in a server room and in a large hall; if a method cannot detect its failure in a small, well-instrumented room, it will not detect it in a large, poorly instrumented one. Small datasets test the fundamental claims; large operational datasets test scale, representativeness and economic value.

23.6 Question-first use of datasets and four experiments

This section turns the catalogue and the capability analysis into a method: start from the question, determine the measurements it needs, check whether a dataset contains them, and design experiments that measure what can be inferred and what cannot. It closes with the form in which the result is stated.

23.6.1 The question before the dataset

The productive order of work is not "here is a dataset; what can be done with it?" but "what question is to be answered?" (can a sensor bias be detected; can the airflow be estimated), followed by "which measurements does it require?" and only then "which dataset contains them?". Starting from the dataset invites the analyses the data make easy, which are usually the compute-side analyses of Section 23.2, and produces results whose physical meaning is unclear. Starting from the question makes the required measurements explicit and makes the answer "this dataset cannot support this question" a legitimate result.

A different objective from benchmarking. Public datasets are often used as training data and benchmarks: a method is trained to predict one recorded variable from the others and scored by its prediction error. In a physical-inference study the datasets serve a different purpose. They are used to test what can and cannot be inferred from a given instrumentation configuration: which balances close, which parameters are identifiable, which hypotheses can be separated. The score is not a prediction error but a set of capability statements, each backed by an experiment.

23.6.2 Four questions and their minimum measurements

Four questions illustrate the method. Table 23.3 lists, for each, the minimum measurements, the relation that uses them, and what the typical public dataset of Section 23.2 offers.

Can airflow be inferred? The minimum is rack power, inlet temperature and outlet temperature, through equation (23.4). Strong validation also needs a measured airflow, to check the inferred one; the server-room dataset of Section 23.2.6 contains both sides of that test.

Can recirculation be diagnosed? The recirculation fraction of Chapter 6 follows from the mixing rule

$$T_{\text{in}} = (1-\alpha)\,T_{\text{supply}} + \alpha\,T_{\text{hot}}, \qquad \alpha = \frac{T_{\text{in}} - T_{\text{supply}}}{T_{\text{hot}} - T_{\text{supply}}}, \tag{23.10}$$

so the minimum is the supply temperature T_{supply}, the hot-aisle temperature T_{hot} and the rack inlet temperature T_{in}, together with the spatial placement of the sensors. With illustrative values $T_{\text{supply}} = 18°\text{C}$, $T_{\text{hot}} = 36°\text{C}$ and $T_{\text{in}} = 21.6°\text{C}$, $\alpha = 3.6/18 = 0.20$. Server power and CPU temperature alone are insufficient: they do not contain T_{supply} or T_{hot}, and a CPU temperature is the end of a heat path, as in Section 23.2.3.

Can chiller degradation be estimated? The minimum is the cooling delivered, the chiller's electrical power, the chilled-water temperatures, the condenser-water temperature T_{CW}, and the outdoor temperature or wet-bulb temperature. Degradation appears as a falling coefficient of performance at fixed conditions, $\text{COP} = \dot{Q}_{\text{evap}}/P_{\text{chiller}}$ in the electrical form of Chapter 7: an illustrative chiller delivering 1500 kW of cooling for 300 kW of electrical power has COP = 5.0; if, at the same water and outdoor temperatures, it later needs 357 kW, COP = 4.2, a 16 % loss. Most public compute datasets cannot help, because they contain none of these variables.

Can thermal risk be estimated? With $P(t)$, $T_{\text{in}}(t)$ and $T_{\text{out}}(t)$, a lumped thermal model of Chapter 8,

$$C_{\text{th}}\,\frac{\mathrm{d}T}{\mathrm{d}t} = P - \dot{Q}_{\text{removed}}, \tag{23.11}$$

can be fitted, and, if it is validated, used to state the probability $\Pr\big(T(t+h) > T_{\max}\big)$ that a temperature exceeds its limit within a horizon h, as in Chapter 17. If the parameters are poorly identifiable, the forward prediction is not reliable. With an illustrative $C_{\text{th}} = 600\,\text{kJ/K}$ and a 10 kW deficit after a cooling loss, the temperature rises at $10/600\,\text{K/s} = 1\,\text{K/min}$ and an 8 K margin lasts 8 minutes; if C_{th} is known only to ±30 %, the time to the limit lies anywhere from 5.6 to 10.4 minutes. The honest result is then a capability statement: "the dataset supports detection but not reliable forward risk prediction".

Table 23.3: Four questions, the minimum measurements each requires, the relation that uses them, and what typical public compute datasets offer.

Question	Minimum measurements	Relation	Typical public data
Airflow	P_{rack}, T_{in}, T_{out}; $\dot{m}_{\text{air}}$ for validation	$\hat{\dot{m}} = P/(c_p \Delta T)$, eq. (23.4)	thermal datasets: estimate; server-room data: estimate and validation
Recirculation	T_{supply}, T_{hot}, T_{in}, sensor placement	mixing rule, eq. (23.10)	power and CPU temperature: insufficient
Chiller degradation	cooling delivered, P_{chiller}, $T_{\text{CHW,s}}$, $T_{\text{CHW,r}}$, T_{CW}, T_{amb} or T_{wb}	$\text{COP} = \dot{Q}_{\text{evap}}/P_{\text{chiller}}$	compute datasets: none of these
Thermal risk	$P(t)$, $T_{\text{in}}(t)$, $T_{\text{out}}(t)$; validated dynamics	eq. (23.11); $\Pr(T(t+h) > T_{\max})$	detection supported; forecast only if parameters identifiable

23.6.3 First experiment: closure of the balance

With a dataset that records server power, temperature rise and measured airflow, such as the server-room dataset of Section 23.2.6, the first experiment asks whether a reduced physical model can consistently reconcile the three. The airflow residual $r_{\dot m}$ of equation (23.5), or the energy residual r_Q of equation (23.7), is computed over the whole record, and each period is classified into one of four classes:

- *balance closes:* the residual is within its expected spread, set by the stated accuracy of the sensors;
- *balance fails:* the residual is persistently outside that spread;
- *parameter drift:* the balance closes only if an effective parameter, such as the ratio of measured to calculated airflow, changes slowly over time;
- *unexplained residual:* a residual with structure that none of the above describes.

The second half of the experiment asks whether the structure of the residual is explained by recirculation (a residual that grows when the inlet temperature rises relative to the supply), by sensor bias (a constant offset in one sensor's terms, the same at all loads), or by model inadequacy (a residual that tracks load or time, such as the lag of Section 23.3.2). The output is a map of where and when the simple model holds.

23.6.4 Second experiment: airflow withheld and interval calibration

The second experiment hides the measured airflow from the inference, infers the posterior density $p(\dot m \mid P, T_{\text{in}}, T_{\text{out}})$ from the remaining variables, and then reveals the measured airflow to compare. The comparison has two parts.

The first part is accuracy, measured by the root-mean-square error over n periods,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\hat{\dot m}_k - \dot m_{\text{meas},k}\right)^2}, \tag{23.12}$$

where $\hat{\dot m}_k$ is the center of the inferred distribution in period k. For four illustrative errors of -0.04, 0.02, 0.06 and $-0.02\,\text{kg/s}$, the mean square is $0.0060/4 = 0.0015$ and the RMSE is $0.039\,\text{kg/s}$.

The second part is uncertainty calibration, which asks whether the method's stated uncertainty is honest. In each period the method gives an interval it says contains the true airflow with 95 % probability. After the measured airflow is revealed, one counts the periods in which it falls inside its interval. If the intervals are honest, about 95 % of the periods are inside. Over 400 periods that means about 380, with a binomial standard deviation of $\sqrt{400 \times 0.95 \times 0.05} \approx 4.4$, so a count between about 371 and 389 is consistent with calibration. A count of 270, about 67 %, means the intervals are too narrow and the method is *overconfident*: it claims more certainty than it has. A count of 399 with wide intervals means it is *underconfident*: correct but needlessly vague. Because the revealed airflow is itself a measurement, the interval compared with it must include the airflow meter's own error: with an inferred standard deviation of $0.05\,\text{kg/s}$ and a meter standard deviation of $0.03\,\text{kg/s}$, the interval for the measured value has a standard deviation of $\sqrt{0.05^2 + 0.03^2} = 0.058\,\text{kg/s}$.

Figure 23.9 draws the calibration curve, computed by a short script for Gaussian errors and synthetic by construction. The horizontal axis is the nominal coverage the method claims for its intervals, from 0 to 1; the vertical axis is the fraction of revealed values that actually fall inside. The diagonal is perfect calibration. The three curves correspond to methods whose stated standard deviation is half, equal to and one and a half times the true spread of their

errors. At the nominal 95 % the half-width method covers only 67 %, while the wide method covers 99.7 %. A calibrated method lies on the diagonal at every level, not only at 95 %.

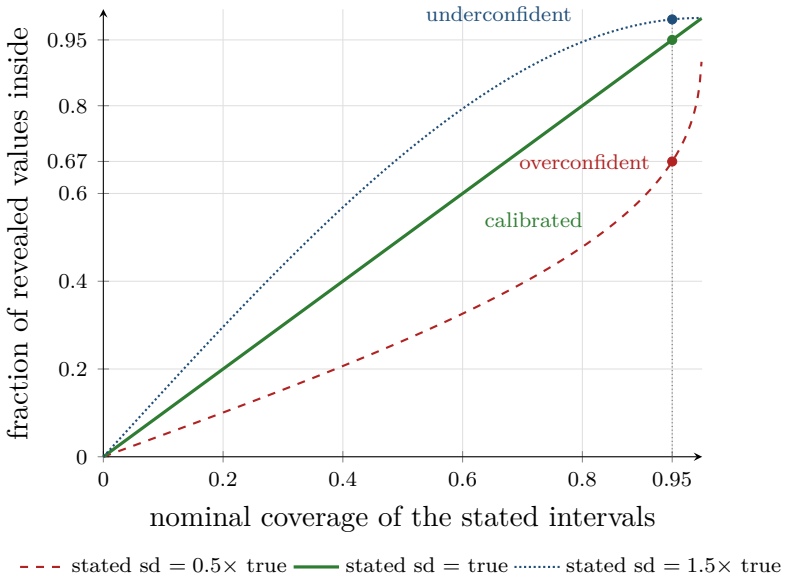


Figure 23.9: Interval calibration for Gaussian prediction errors; synthetic, computed by the chapter's data script and not taken from any dataset. Empirical coverage $2\Phi(z_q\, s_{\text{stated}}/s_{\text{true}}) - 1$ against nominal coverage q, with $z_q = \Phi^{-1}((1+q)/2)$, for stated standard deviations of 0.5, 1 and 1.5 times the true one. At the nominal 95 % level the coverages are 0.673, 0.950 and 0.997.

23.6.5 Third and fourth experiments: bias injection and removal

Third experiment. The third experiment injects a bias of +2°C into one temperature sensor, say the outlet sensor, and asks the inference to compare two hypotheses: H_1, a sensor bias, and H_2, an airflow change. With the illustrative 10 kW rack, a true 12 K rise and a measured airflow of 0.833 kg/s, the biased record shows a 14 K rise. Under H_1 the measured airflow stays at 0.833 kg/s, and the energy residual becomes $r_Q = 10 - 0.833 \times 1.0 \times 14 = -1.67$ kW. Under H_2 there is no bias and the balance closes only if the airflow fell to $10/14 = 0.714$ kg/s, 14 % less, which the airflow meter would show. The measured airflow therefore separates the two: an unchanged meter reading with a negative residual favors H_1; a lower meter reading with a closed balance favors H_2.

The experiment is then repeated over a range of bias magnitudes to find the detection limit. A bias b on the outlet sensor produces $r_Q = -\dot{m}_{\text{air}} c_{p,\text{air}}\, b = -0.833\, b$ kW: -0.42 kW at 0.5 K, -0.83 kW at 1 K, -1.67 kW at 2 K. Random noise in the residual shrinks with averaging over many periods, but a persistent error in the other sensors does not. If the airflow meter has a persistent error of 5 %, it alone produces a residual of up to $0.05 \times 10 = 0.5$ kW, the same as a temperature bias of $0.5/0.833 = 0.6$ K. Below about 0.6 K the two hypotheses are indistinguishable whatever the record length; that floor, set by the accuracy of the other sensors, is the detection limit the experiment measures.

Fourth experiment. The fourth experiment removes the airflow from the dataset and asks whether the same fault can still be diagnosed. Without the meter, H_1 and H_2 predict the same power, the same inlet temperature and the same outlet temperature: a 14 K rise at 10 kW is a biased sensor on a normal flow or a correct sensor on a flow of 0.714 kg/s. The diagnosis is not identifiable. The correct output is not a guess but the statement "diagnosis not identifiable;

measure airflow", which links three steps of the book: inference, then the identifiability limit of Chapter 15, then the next-best measurement of Chapter 19. Comparing the third and fourth experiments measures the value of the airflow meter directly: it is the difference between a diagnosis and an ambiguity.

23.6.6 Capability as an explicit output

The four experiments end in the same kind of result: a statement, per question, of what the dataset supports and why. Table 23.4 gives such a statement for an illustrative instrumentation configuration with server power, inlet and outlet temperatures, measured air speed and known sensor geometry, and with no chiller or electrical-backup telemetry. Each row is a capability, a verdict and the reason, which is always a measurement present or absent. Heat-load estimation is supported because power is measured; airflow estimation is supported, conditionally, because power and the temperature difference are available; airflow validation is supported because air speed is measured; sensor-bias detection is possible because the physics is redundant; chiller diagnosis and UPS reliability are not supported because the relevant telemetry is absent; thermal-risk forecasting is conditional on validated dynamics; recirculation diagnosis is conditional on the spatial geometry of the sensors.

Table 23.4: A capability statement as the explicit output of an analysis, for an illustrative configuration with server power, inlet and outlet temperatures, measured air speed and sensor geometry, and no chiller or electrical-backup telemetry.

Capability	Verdict	Reason
Heat-load estimation	yes	power measured
Airflow estimation	yes, conditional	P and ΔT available; assumptions of Section 23.2.5
Airflow validation	yes	air speed measured
Sensor-bias detection	possibly	redundant physics, limited by the other sensors' accuracy
Chiller diagnosis	no	no chiller telemetry
UPS reliability	no	no electrical backup data
Thermal-risk forecast	conditional	validated dynamics required
Recirculation diagnosis	conditional	spatial sensor geometry required

A table of this kind belongs in every report of an inference study, before any result. It answers the question that matters more than whether a public dataset exists: what physical questions does each dataset contain enough independent information to answer, and which remain unsupported because the measurements are missing. The contrast with a machine-learning workflow is one of question. A machine-learning workflow asks whether a target y can be predicted from the features. A physical-inference workflow asks whether the measurement system contains enough physical information to justify the inference at all.

Principle 23.3. *The capability of a dataset is a result, not an assumption: for each question, state whether the measurements support it, conditionally support it or cannot support it, and name the measurement that decides the verdict.*

23.7 What to retain

Definition 23.1 (Capability of a dataset)**.** The capability of a dataset is the set of physical questions its measurements contain enough independent information to answer; it is fixed by

which balances its synchronized, adequately sampled variables allow to close, it grows with each measurement that completes a balance, and it is reported per question, with the missing measurement named wherever a question is unsupported.

Table 23.5 lists the chapter's objects and the subsection that defines each.

Table 23.5: The objects of the chapter and where each is defined.

Object	Expression	Role	Defined in
Four data classes	Figure 23.1	workload, power, thermal, integrated	23.1.2
Dataset template	Table 23.1	contains, supports, cannot support	23.2.1
GPU identifiability limit	eq. (23.3)	cause not separable without T_{in}	23.2.3
Airflow estimate	$\hat{\dot{m}} = P/(c_p \Delta T)$, eq. (23.4)	calculated airflow; five assumptions	23.2.5
Airflow residual	$r_{\dot{m}}$, eq. (23.5)	test of the balance	23.2.7
Ideal dataset	Figure 23.4	seven synchronized groups	23.3.1
Aliased frequency	f_{alias}, eq. (23.6)	false slow variation	23.3.3
Capability map	Table 23.2	capability by instrumentation	23.4.1
Energy residual	r_Q, eq. (23.7)	balance with measured airflow	23.4.2
Flow resistance	$K = \Delta p/\dot{V}^2$	blockage as rising K	23.4.3
Coil balance	$\dot{Q}_w$ against $\dot{Q}_{\text{air}}$	redundant energy check	23.4.5
Synthetic fault	b_T, $\dot{m}' = 0.7\,\dot{m}$	injected ground truth	23.5.2
Validation stages	Figure 23.8	synthetic to operational	23.5.4
Interval calibration	Figure 23.9	honesty of stated uncertainty	23.6.4
Capability statement	Table 23.4	explicit output per question	23.6.6

23.8 Exercises

Exercise 23.1. A rack draws 8 kW with $T_{\text{in}} = 22°\text{C}$ and $T_{\text{out}} = 34°\text{C}$; take $c_p = 1.0\,\text{kJ/(kg K)}$. (a) Compute the airflow estimate of equation (23.4). (b) Each temperature sensor may be off by up to 0.5 K; compute the range of the estimate in the worst case. (c) A measured mean air speed of 1.4 m/s over a 0.45 m^2 face at $\rho = 1.2\,\text{kg/m}^3$ is available. Compute the measured airflow and the residual $r_{\dot{m}}$ of equation (23.5). (d) Is the residual explained by the temperature tolerance alone? Name two of the five assumptions of Section 23.2.5 that could explain it, and the direction of the error each would cause.

Exercise 23.2. A chilled-water pump's speed oscillates with a period of 130 s and is recorded every 120 s. (a) Compute the Nyquist frequency of the record and the apparent period of the oscillation from equation (23.6). (b) A 30 s event occurs at a random time; compute the probability that the two-minute record contains at least one reading during it. (c) What sampling interval would show both the oscillation and the event with at least three readings per oscillation period and per event?

Exercise 23.3. At a group of coils, the air side gives $\dot{m}_{\text{air}} = 30\,\text{kg/s}$ with a 11 K temperature drop across the coils, and the water side gives $\dot{m}_w = 15\,\text{kg/s}$ with $T_{\text{CHW,s}} = 12°\text{C}$ and

$T_{\mathrm{CHW,r}} = 16.5°\mathrm{C}$; take $c_{p,\mathrm{air}} = 1.0\,\mathrm{kJ/(kg\,K)}$ and $c_{p,w} = 4.18\,\mathrm{kJ/(kg\,K)}$. (a) Compute $\dot{Q}_{\mathrm{air}}$ and $\dot{Q}_w$, the gap, and the gap as a fraction of the air-side heat. (b) What error in the water temperature difference alone would close the gap? (c) What error in the air mass flow alone would close it? (d) Which of the six explanations of Section 23.4.6 should be tested first, and how does repeating the comparison over steady periods help?

Exercise 23.4. An inference method withholds the measured airflow and states a 95 % interval in each of 250 periods. (a) How many periods should contain the measured airflow if the method is calibrated, and what is the binomial standard deviation of that count? (b) The count is 214. Compute the observed coverage and say whether the method is overconfident or underconfident. (c) Assuming Gaussian errors, use the relation in the caption of Figure 23.9 to estimate the ratio of the stated to the true standard deviation ($\Phi(1.46) \approx 0.928$). (d) The airflow meter has a standard deviation of 0.03 kg/s and the method states 0.04 kg/s for the inferred airflow; what standard deviation should the interval compared with the meter reading use?

23.9 Connection to the TERRA framework

A dataset enters the TERRA framework as a set of measurement equations attached to a graph, and its capability is a property of the inference on that graph.

- **Each measured variable is a measurement equation on a node or an edge.** Rack power is the electrical flow into a rack node, T_{in} and T_{out} are temperatures at the aisle and rack nodes, a measured airflow is the mass flow on the rack's air edge, and $\dot{Q}_w$ is the heat on the coil's water edge. The seven groups of Figure 23.4 are measurement equations on the three coupled networks, electrical, air and water mass, and thermal energy; the topology G is the graph itself, and an uncertain one where the dataset does not document it.

- **A capability map is the identifiability of the graph's unknowns.** With P_{rack}, T_{in} and T_{out} only, the rack airflow is a latent edge flow inferred through the rack's energy balance, and the filter resistance and fan state are not separately identifiable; each column of Table 23.2 adds measurement equations that make further unknowns identifiable.

- **Redundant balances are residuals.** The energy residual r_Q at a rack node and the comparison of $\dot{Q}_{\mathrm{air}}$ with $\dot{Q}_w$ at a coil node are balances that should close; a sensor bias b is a term in one measurement equation, and an injected fault is a known bias or a known scaling of an edge flow, such as $\dot{m}' = 0.7\,\dot{m}$.

- **Withheld measurements test the posterior.** The second experiment is the framework's posterior $p(\dot{m} \mid P, T_{\mathrm{in}}, T_{\mathrm{out}})$ for a latent edge flow, scored against the withheld meter; interval calibration (Figure 23.9) checks that the posterior's uncertainty is honest.

- **Hypotheses and the next measurement.** H_1, a bias in the outlet measurement equation, and H_2, a reduced flow on the rack edge, receive posterior probabilities $\Pr(H_i \mid y)$; when the airflow measurement is removed they become equal, and the framework's output is the identifiability limit together with the measurement that would resolve it.

In summary, public data are rich in workload, power and temperature and poor in airflow, pressure, water-side and topology data; a dataset's value lies in the balances its synchronized, adequately sampled measurements allow to close; each added measurement extends what can be inferred in a way that can be tabulated in advance; validation proceeds through synthetic models, public data, controlled experiments and operation because no single setting is both realistic and labeled; and the capability of a dataset is an explicit output of the analysis, stated

per question before any inference result. The next chapter gathers the book's threads into one synthesis of the data center as a physical system observed, modeled and inferred (Chapter 24).

Chapter 24

TERRA for Data Centers

24.1 Existing tools and the inference gap

The preceding chapters followed energy through a data center, from the utility connection to the heat rejected outdoors, and then asked how the facility measures that chain, how it fails, how a fault is diagnosed, how capacity and risk are stated, and how the next measurement and the next action are chosen. This last chapter places one framework against that material and asks:

where does a probabilistic physical-inference framework add something distinctive in a data center, rather than duplicating tools that already exist?

The framework is TERRA, a graph-based framework that models a physical system as conservation laws on nodes with flows on edges, with probabilistic inference of the states and parameters that the measurements do not give directly. The answer argued below is that it does not add something everywhere. It adds something where a problem combines four properties: physics that constrains the answer, uncertainty in the state or the parameters, competing explanations of the same observations, and an operational consequence that depends on which explanation is true.

The chapter assesses fit. It maps the material of the book onto eight use cases, states where the framework is not differentiated, describes the form its outputs would take, and sets out a path by which it could be introduced. None of the use cases is reported here as a demonstrated result on a real facility. Every number in the chapter is illustrative, and every statement of what TERRA would compute, rank or recommend is a proposal. The commercial statements (which use case is stronger, what a first deliverable should be, what should not be claimed) are the chapter's own assessment and are stated as such.

Jargon: CFD, MPC, digital twin, inference layer, epistemic uncertainty, risk-qualified capacity, instrumentation audit, maturity ladder

CFD, computational fluid dynamics, solves the equations of fluid motion and heat transport on a fine spatial mesh of a room or a piece of equipment. *MPC*, model predictive control, chooses control actions by optimizing a model's predicted trajectory over a horizon and applying the first step. A *digital twin* is a model of a specific facility that is kept up to date with that facility's data; it need not be a CFD model. An *inference layer* is software that sits between data and decisions and turns measurements into statements about unmeasured states, parameters and causes, each with a probability. *Epistemic uncertainty* is uncertainty that comes from incomplete knowledge (an unmeasured airflow, an unknown sensor bias) and can in principle be reduced by more information, as distinct from irreducible randomness. A *risk-qualified capacity* is a capacity stated together with the probability that it is not available and with an explicit evaluation of the uncertainty and the model behind it. An *instrumentation audit* is a structured review of what a facility measures, what it does not, and which operational questions its measurements can answer. A *maturity ladder* is an ordered list of capabilities, each of which requires the ones below it. BMS and DCIM, the building management system and the data center infrastructure management software, are defined in the vocabulary at the front of the book.

Key idea

The physics of a data center is well understood; what is rarely known is its current state. A physical-inference framework is distinctive where a decision depends on states, parameters or causes that the telemetry does not give directly, and where the honest answer is a probability, a ranked set of explanations, or a statement of what must be measured before the question can be answered.

24.1.1 The tool landscape

A data center already runs on four families of software, and each answers its own kind of question well. Table 24.1 lists them with what each provides, the question it answers, and the kind of question it is not built to answer. The building management system and the data center infrastructure management software collect telemetry, raise alarms, record equipment state, draw trends and dashboards, and keep the capacity inventory, the record of what is installed and how much of it is allocated (Chapter 9). CFD resolves airflow and temperature in detail and is used to find hot spots, to design rooms and to evaluate a layout change before it is made (Chapter 22). Optimization and model predictive control choose setpoints, minimize energy, schedule workload and run the control loops (Chapters 10 and 20). Machine learning forecasts load, detects anomalies, recognizes patterns and supports predictive maintenance.

The last column of the table is where a physical-inference framework would work. It should not be another dashboard, because the dashboard already exists and displays the readings well; and it should not be another CFD solver, because CFD already resolves the field in detail when its boundary conditions are known. The space left is the question that each existing tool assumes has already been answered: what is the current state of the physical system, given telemetry that is sparse, possibly biased and possibly contradictory?

24.1.2 Questions of inference

The questions that the existing tools leave open have a common form. Four of them recur in operation:

- "The measurements do not agree; which explanation is plausible?"

Table 24.1: The existing software of a data center and the kind of question each family answers.

Tool family	Provides	Answers	Not built to answer
BMS, DCIM	telemetry, alarms, equipment state, trends, dashboards, capacity inventory	what do the sensors read; what is installed	whether a reading is credible; why it changed
CFD	detailed airflow and temperature fields, hot spots, layout studies	what happens in this geometry with these boundary conditions	which boundary conditions hold today
Optimization, MPC	setpoints, energy minimization, workload scheduling, control	which action is best under a given model	whether the model's parameters are known well enough to act on
Machine learning	forecasts, anomaly detection, pattern recognition, predictive maintenance	what usually follows this pattern	which physical mechanism caused an anomaly never seen before

- "Can this capacity estimate be trusted?"
- "Is this a cooling problem or a bad sensor?"
- "What additional measurement would resolve the ambiguity?"

None of these asks for a new reading or for a finer simulation. Each asks for a conclusion about something that is not measured, a cause, a bias, a capacity, the value of a measurement, drawn from measurements that are. In the vocabulary of this book they are inference questions, as distinct from measurement questions (what does the sensor read) and calculation questions (what follows from the readings by a known relation). Chapters 14 to 19 developed the methods behind them: state estimation, identifiability, probabilistic diagnosis, capacity under risk, what-if analysis and the choice of the next measurement. The formal object common to all of them is a posterior distribution, $p(x, \theta \mid y)$ over states and parameters or $\Pr(H_i \mid y)$ over hypotheses, computed from a physical model and the measurements y.

24.1.3 A physical-inference layer

The role proposed for TERRA in a data center follows from the two previous subsections. It would be a probabilistic physical-inference layer for operational decision support, placed between the telemetry and the models on one side and the operator's decisions on the other. Its inputs are sensors, models and context (the maintenance state, the topology G, the weather); its outputs are an estimate of the state, a diagnosis, a risk and a recommended next action.

Figure 24.1 draws this placement as a stack of four layers. The bottom layer is the telemetry that the facility already collects. The second is the set of physical models that already exist or can be built, from equipment curves to reduced-order thermal networks to CFD. The third, drawn in green, is the inference layer, with its seven functions as boxes: model credibility, state estimation, diagnosis, uncertainty, risk, the next measurement and what-if analysis. The top layer is where decisions are taken, by an operator, by a supervisory optimizer or by the control system. The arrows between layers carry what each layer passes upward; the dashed arrow on the left carries the recommended measurement back down to the telemetry. The figure shows that the inference layer replaces none of the others: it consumes the two below and serves the one above.

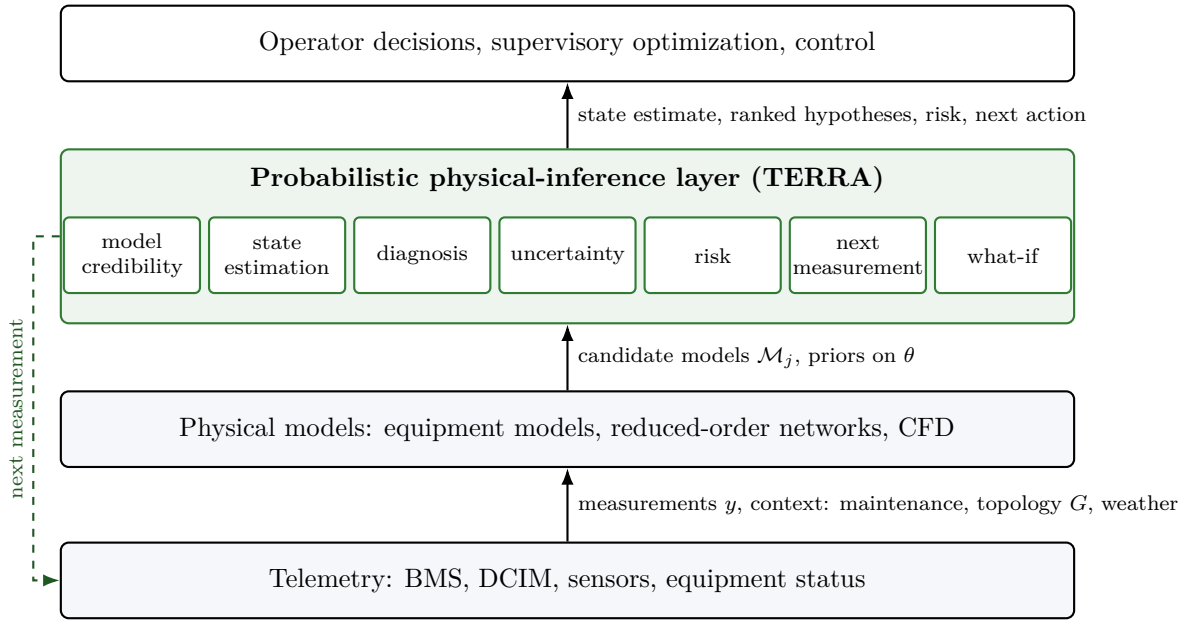


Figure 24.1: The proposed place of a physical-inference layer in the software stack of a data center. Each layer passes the labeled quantities upward; the dashed arrow returns the recommended measurement to the telemetry layer. The inference layer consumes the layers below it and serves the one above; it replaces none of them.

24.1.4 Four properties of a distinctive problem

The questions of Section 24.1.2 share four properties, and a problem that has all four is the kind on which a physical-inference layer would be distinctive.

Physics. The answer is constrained by conservation laws and equipment relations: the heat a rack generates equals its electrical power (Chapter 5), the heat the air carries is $\dot{Q} = \dot{m}\, c_p\, \Delta T$, the inlet temperature of a rack follows the recirculation mixing rule (Chapter 6), and a room warms no faster than its thermal capacitance allows (Chapter 8). Without physics, a statistical model can only say what usually happens; with it, a model can say what is possible.

Uncertainty. Some of the quantities the answer depends on are not measured: rack airflow, the recirculation fraction α, the conductance UA of a fouled coil, the bias b of a sensor, the topology G of the power paths. They must be inferred, and their uncertainty must be carried into the conclusion (Chapters 14 and 15).

Competing explanations. More than one mechanism produces the same symptom. A rising inlet temperature can be caused by load, airflow, recirculation, chilled water or the sensor itself, and the conclusion is a ranking of hypotheses rather than a single value (Chapters 13 and 16).

Operational consequence. The decision depends on which explanation is true: migrate workload or dispatch a technician, accept the new rack or decline it, call the standby chiller or wait. If every explanation led to the same action, the ranking would have no value (Chapter 19).

A problem without physics is a statistics problem; without uncertainty it is a simulation or an optimization problem; without competing explanations it is an estimation problem; without consequence it is an academic one. The use cases of the next two sections are chosen because each has all four.

Principle 24.1. *A physical-inference layer is distinctive only where physics, uncertainty, competing explanations and operational consequence occur together. Where any one of them is absent, an existing tool usually suffices.*

24.2 Use cases in diagnosis and measurement

This section describes four of the eight proposed use cases, those that concern what the telemetry says about the present state: whether a reading can be trusted, which mechanism explains a symptom, which measurement would resolve an ambiguity, and whether a model can be trusted. Each is stated as what TERRA would compute; none is reported as a result on a real facility.

24.2.1 Use case 1: sensor trust

An inlet sensor on rack 18 reads $T_{\text{in}} = 34°\text{C}$, above its high-temperature alarm threshold, and the alarm fires. The operator's question is not whether the sensor reads 34°C, which the alarm already states, but whether the rack is actually hot. The reading is measured; the hot spot is a hypothesis. Two hypotheses compete: a real hot spot, H_{hot}, and a sensor bias, H_{bias}, in which the sensor reads high while the air is at its normal temperature.

The physical model provides the evidence that separates them. A real hot spot of this size would be visible in quantities other than the one sensor: in the rack's electrical power, if load caused it; in the inlet temperatures of neighboring racks, which share the same aisle air; in the airflow and the state of the cooling units; in the servers' own internal temperature sensors, which would rise with the inlet; and in the time course of the reading itself, since a real inlet temperature cannot rise faster than the air volume and thermal capacitance of the aisle allow (the time constant $\tau = R_{\text{th}}C_{\text{th}}$ of Chapter 8). A sensor bias predicts the opposite: the other quantities stay normal, and the reading may step abruptly.

Suppose the rack power is unchanged, the neighboring inlets read normal, and the servers' internal temperatures are normal. With illustrative likelihoods for each piece of evidence under each hypothesis, and the three pieces taken as conditionally independent given the hypothesis, Bayes' rule in odds form (Chapter 16) gives

$$\frac{\Pr(H_{\text{hot}} \mid y)}{\Pr(H_{\text{bias}} \mid y)} = \frac{\Pr(H_{\text{hot}})}{\Pr(H_{\text{bias}})} \times \frac{0.1}{0.8} \times \frac{0.4}{0.9} \times \frac{0.6}{0.9} = 1 \times 0.125 \times 0.444 \times 0.667 = 0.037, \qquad (24.1)$$

where the three ratios are, in order, the probability of normal internal server temperatures, of normal neighboring inlets and of unchanged rack power under each hypothesis, and the prior odds are taken as even. The posterior probabilities are

$$\Pr(H_{\text{hot}} \mid y) = \frac{0.037}{1 + 0.037} = 0.036, \qquad \Pr(H_{\text{bias}} \mid y) = 0.964.$$

The alarm said the rack is hot. The physical evidence says that it is probably not: the reading is very likely a sensor fault. The proposed output of the use case is this pair of probabilities, $\Pr(\text{real hot spot} \mid y)$ against $\Pr(\text{sensor bias} \mid y)$, with the evidence behind each.

Operational value of measurement credibility. Both errors cost something. A false alarm treated as real triggers unnecessary workload migration, extra cooling, a technician dispatch and disruption to the services on the rack. A real hot spot dismissed as a sensor fault damages availability, the quantity the facility exists to deliver. The credibility of a measurement therefore has operational value, and it is the lowest rung of the capabilities described in Section 24.6.4: before any state, risk or capacity is inferred from a sensor, the sensor itself must be credible.

24.2.2 Use case 2: competing cooling diagnoses

The inlet temperature of a rack rises. Five hypotheses explain the symptom:

- H_1: the IT load of the rack has increased;
- H_2: the airflow delivered to the rack has fallen;
- H_3: hot exhaust air recirculates into the rack inlet;
- H_4: the chilled water reaching the air handlers is warmer;
- H_5: the inlet sensor is biased.

An alarm compares the reading with a limit, $T > T_{\max}$, and reports that the limit is exceeded. The proposed inference reports the posterior $\Pr(H_i \mid y)$ for each hypothesis, the quantity developed in Chapter 16. The two outputs answer different questions: the alarm says that something is wrong; the posterior says what is probably wrong and how sure that conclusion is.

The posterior is only as discriminating as the evidence. Each hypothesis leaves a different signature in the quantities around the rack, and Table 24.2 lists them. Its rows are the five hypotheses; its columns are the rack inlet temperature and four other quantities a facility may or may not measure. The first column is the same in every row: every hypothesis raises the inlet reading. The other columns differ: only H_1 raises the rack power, only H_2 lowers the rack airflow, only H_4 raises the supply air temperature, H_3 shows in the neighboring hot-aisle temperature, and H_5 shows in nothing but the one sensor.

Table 24.2: Signatures of five explanations of a rising rack inlet temperature. "up", "down" and "normal" describe the expected change of each quantity under each hypothesis.

Hypothesis	Inlet reading	Rack power	Rack airflow	Supply air temperature	Neighboring hot aisle
H_1 higher IT load	up	up	normal	normal	up (local)
H_2 lower airflow	up	normal	down	normal	normal
H_3 recirculation	up	normal	normal	normal	up
H_4 warmer chilled water	up	normal	normal	up	up (room-wide)
H_5 sensor bias	up	normal	normal	normal	normal

24.2.3 Diagnosis only where the data support it

If the only measurement is the rack inlet temperature, the first column of Table 24.2 is all the evidence there is, and it is the same under every hypothesis. The likelihoods are then equal, $p(y \mid H_i) = p(y \mid H_j)$ for every pair, and Bayes' rule returns the prior unchanged:

$$\Pr(H_i \mid y) = \frac{p(y \mid H_i)\,\Pr(H_i)}{\sum_j p(y \mid H_j)\,\Pr(H_j)} = \Pr(H_i) \quad \text{when } p(y \mid H_i) \text{ is the same for all } i. \tag{24.2}$$

The five explanations cannot be separated, and any ranking reported in that case is a ranking of priors presented as a diagnosis. This is the non-identifiability of Chapter 15 applied to hypotheses rather than to parameters.

The proposed output in that case is a statement of restraint with guidance attached:

> "Current instrumentation does not distinguish these mechanisms. An airflow measurement would most strongly separate H_2 from H_3."

The second sentence follows from Table 24.2: the rack airflow column is the one in which H_2 and H_3 differ, and these two are the hypotheses that the cooling side cannot otherwise tell apart, since neither changes the supply air or the rack power. Restraint with next-step guidance is the distinctive output here. A diagnostic tool that always names a cause will name one in this case too, and it will be right only as often as the prior is.

> **Principle 24.2.** *An inference is reported only as far as the measurements support it. When the evidence cannot separate the hypotheses, the output says so and names the measurement that would separate them.*

24.2.4 Use case 5: the next-best measurement

Suppose the temperatures available leave two hypotheses almost tied: recirculation at $\Pr(H_3 \mid y) = 0.46$ and reduced airflow at $\Pr(H_2 \mid y) = 0.43$, with the remaining 0.11 shared by high IT load (0.03, made unlikely by the rack power meter), warm chilled water (0.04) and sensor bias (0.04). Forcing a diagnosis would name recirculation, and would be wrong with probability $1 - 0.46 = 0.54$. The proposed output is instead the measurement that would change the decision most, with its reason: "measure the rack airflow; it separates recirculation from reduced airflow, which call for different actions (containment work against fan or filter maintenance)."

The calculation behind the recommendation is the expected information gain of Chapter 19. Let the airflow measurement return "low" or "normal", with illustrative probabilities of a low reading 0.10, 0.90, 0.15, 0.10 and 0.10 under H_1 to H_5. The probability of a low reading is

$$\Pr(\text{low}) = 0.03(0.10) + 0.43(0.90) + 0.46(0.15) + 0.04(0.10) + 0.04(0.10) = 0.467.$$

After a low reading the posterior of reduced airflow becomes $0.43 \times 0.90/0.467 = 0.83$ and that of recirculation $0.46 \times 0.15/0.467 = 0.15$; after a normal reading, with probability 0.533, recirculation rises to $0.46 \times 0.85/0.533 = 0.73$ and reduced airflow falls to $0.43 \times 0.10/0.533 = 0.08$. The entropy of the diagnosis, $\mathcal{H} = -\sum_i \Pr(H_i \mid y) \log_2 \Pr(H_i \mid y)$, is 1.56 bits before the measurement, 0.80 bits after a low reading and 1.36 bits after a normal one, so the expected information gain is

$$\text{IG} = 1.56 - (0.467 \times 0.80 + 0.533 \times 1.36) = 1.56 - 1.10 = 0.46 \text{ bit}.$$

Figure 24.2 draws the three posteriors. The horizontal axis lists the five hypotheses; the vertical axis is the posterior probability. The gray bars are the posterior before the airflow measurement, the blue bars the posterior after a low reading and the green bars the posterior after a normal reading. Before the measurement two bars of nearly equal height dominate; after either outcome one bar dominates. Whichever value the measurement returns, the tie is broken, and that is why it is the one recommended.

24.2.5 Use case 6: model credibility

A digital twin of a hall is reported to "fit the historical temperatures well". A good fit is necessary and far from sufficient. The proposed use case is an evaluation of the model's credibility, which asks six questions:

1. Are the parameters identifiable, or does the fit leave directions in parameter space along which the data are silent (Chapter 15)?
2. Are the fitted parameters physically plausible, for example a coil conductance within a reasonable range of its design value?

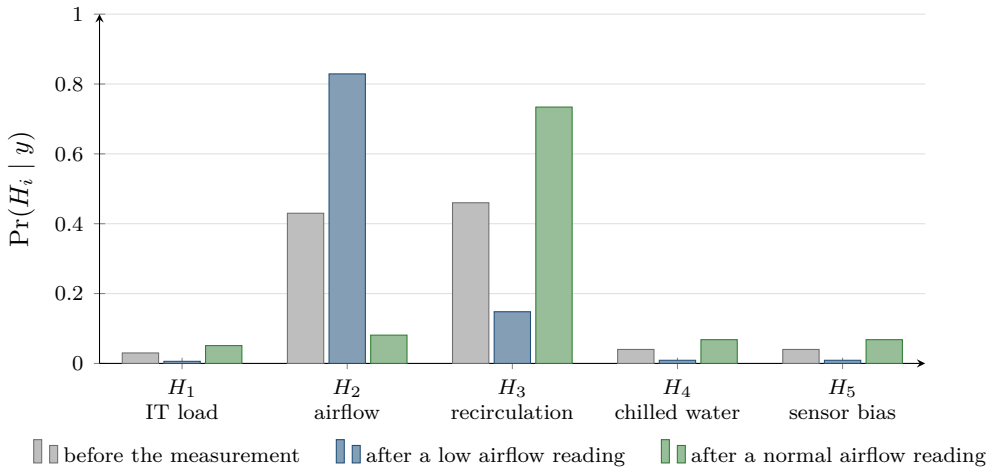


Figure 24.2: Posterior over five explanations of a rising inlet temperature before and after a rack airflow measurement, illustrative. Prior to the measurement: 0.03, 0.43, 0.46, 0.04, 0.04; probability of a low reading under H_1 to H_5: 0.10, 0.90, 0.15, 0.10, 0.10. Either outcome breaks the tie between H_2 and H_3; the expected information gain is 0.46 bit.

3. Does the model close the energy balances: IT power against the air-side heat, the air side against the water side, the evaporator against the condenser (Chapters 5 and 7)?

4. Is the model valid under interventions, that is, does it predict the response to a setpoint change or an equipment outage, not only the routine operation it was fitted to?

5. Is the fit prior-driven, with the parameters held in place by their priors rather than by the data?

6. Are there important residuals the model does not explain, such as a daily pattern in the error that follows a control schedule?

The third question can be answered with a few numbers. Suppose the twin fits the inlet temperatures to within 0.3 K and implies an air-side heat removal of $\dot{m}_{\text{air}}\, c_{p,\text{air}}\, \Delta T = 110\,\text{kg/s} \times 1.0\,\text{kJ/(kg K)} \times 10\,\text{K} = 1100\,\text{kW}$ for the hall, while the measured chilled-water side gives $\dot{m}_w\, c_{p,w}\, \Delta T = 44\,\text{kg/s} \times 4.18\,\text{kJ/(kg K)} \times 5\,\text{K} = 920\,\text{kW}$ (illustrative values). The two sides of the same coil disagree by 180 kW, about 16 % of the air-side figure. The temperature fit is good and the model is not credible: some parameter, probably the inferred airflow, has absorbed an error that the balance exposes.

Digital twins in operation. As digital twins become common in data centers (Chapter 22), the operator's need moves from having a model to knowing when to trust its answer. The four properties that decide it are adequacy for the question asked, identifiability of the parameters the answer depends on, the uncertainty of the answer, and physical consistency with the conservation laws. A credibility evaluation that reports all four is the proposed output of this use case, and it is a precondition for every use case that follows.

24.3 Use cases in capacity, risk and flexibility

This section describes the four use cases that look forward in time: how much load can be added, what a contingency would do under today's conditions, what a proposed change would do with its uncertainty stated, and how much grid flexibility can be offered safely. Each turns an inferred state into a probability of a violation.

24.3.1 Use case 3: risk-qualified capacity

The capacity question is usually asked as "how many more megawatts can the facility take?" The question the proposed use case answers instead is "how much additional IT load can be supported while the probability of a thermal or electrical violation stays acceptable?", the chance constraint of Chapter 17:

$$C_\epsilon = \max\ L \quad \text{subject to} \quad \Pr(\text{thermal or electrical violation} \mid L) < \epsilon, \tag{24.3}$$

where L is the additional IT load, ϵ the risk tolerance and the probability is taken over the posterior of the inferred quantities (airflow, equipment state, sensor biases) and the forecast of the load. With $\epsilon = 0.05$ the result is written C_{95}: the additional load that can be carried with 95 % confidence that no limit is violated.

Commercial meaning. Both errors in a capacity statement are expensive. Over-conservative operation, a rating set by a fixed allowance far below the true limit, strands megawatts of transformers, UPS modules and chillers that were paid for and cannot be sold. Over-aggressive operation, a rating that takes nominal parameter values at face value, raises the probability of a thermal or electrical outage. The value of a risk-qualified capacity is that it states how much of the installed capacity is genuinely usable, at a stated probability, with the reasons for the uncertainty attached.

24.3.2 Attributed capacity uncertainty

A risk-qualified capacity becomes actionable when its uncertainty is attributed to its sources. An illustrative statement reads: $C_{95} = 6.2\,\text{MW}$ of additional IT capacity, with the uncertainty coming from the airflow distribution (45 %), the cooling-plant state (30 %), the rack load forecast (15 %) and the temperature sensors (10 %). The conclusion drawn from it is that to tighten the rating, airflow knowledge should be improved first.

The arithmetic behind that conclusion shows what an attribution means. Take the capacity as normal, $C \sim \mathcal{N}(\mu_C, \sigma_C^2)$, with an illustrative $\sigma_C = 0.8\,\text{MW}$. The 95 % capacity is the value exceeded with probability 0.95,

$$C_{95} = \mu_C - z_{0.95}\,\sigma_C, \qquad z_{0.95} = 1.645,$$

so $\mu_C = 6.2 + 1.645 \times 0.8 = 7.52\,\text{MW}$. Read the percentages as shares s_k of the variance, $\sigma_C^2 = \sum_k s_k\,\sigma_C^2$ with $\sum_k s_k = 1$, which holds when the sources contribute independently. Removing source k entirely leaves $\sigma_C\sqrt{1 - s_k}$ and

$$C_{95}^{(-k)} = \mu_C - z_{0.95}\,\sigma_C\sqrt{1 - s_k}. \tag{24.4}$$

For the airflow, $s = 0.45$: $\sigma = 0.8\sqrt{0.55} = 0.593\,\text{MW}$ and $C_{95} = 7.52 - 1.645 \times 0.593 = 6.54\,\text{MW}$, a gain of 0.34 MW. For the temperature sensors, $s = 0.10$: $\sigma = 0.759\,\text{MW}$ and $C_{95} = 6.27\,\text{MW}$, a gain of 0.07 MW. The airflow is worth five times as much as the sensors, although the sensors are the more obvious thing to recalibrate.

Figure 24.3 draws the four sources. Each row is one source; the bar is its share of the capacity variance, and the column on the right is the 95 % capacity that would result if that source of uncertainty were removed, from equation (24.4). The ranking of the bars and the ranking of the gains are the same, but the gains fall faster than the shares, because a standard deviation is the square root of a variance: removing 10 % of the variance removes only about 5 % of the standard deviation. Removing all uncertainty would not raise C_{95} to μ_C in practice, since a measurement also moves the mean, up or down; the figure shows the effect on the spread alone.

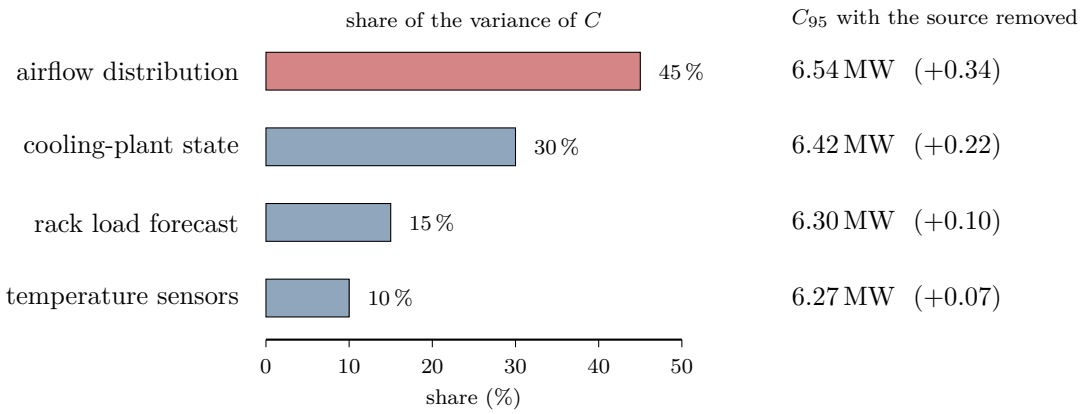


Figure 24.3: Attribution of the uncertainty of an illustrative risk-qualified capacity, $C_{95} = 6.2\,\mathrm{MW}$ with $\sigma_C = 0.8\,\mathrm{MW}$ and $\mu_C = 7.52\,\mathrm{MW}$. Bars: share s_k of the variance of the capacity (the largest in red). Right column: $C_{95}^{(-k)} = \mu_C - 1.645\,\sigma_C\sqrt{1 - s_k}$, equation (24.4), and its gain over 6.2 MW.

24.3.3 Use case 4: contingency risk

A redundancy label such as N+1 says that the design survives the loss of any one unit (Chapter 12). The proposed use case asks a sharper question: given today's load, today's weather, the present condition of each unit, the maintenance state and the uncertainty in all of these, what happens if chiller 3 fails now? The answer has two parts: the probability that a temperature limit is exceeded given the contingency c_3, $\Pr(T > T_{\max} \mid c_3)$, and the time to the limit, t_{limit}, the time the operator has to act (Chapter 18).

An illustrative calculation shows how the two combine. After the loss of chiller 3, the model predicts that the hottest inlet, now at 29°C, rises at $\dot{T} = 0.3\,\mathrm{K/min}$ toward $T_{\max} = 35°\mathrm{C}$, so

$$t_{\mathrm{limit}} = \frac{T_{\max} - T}{\dot{T}} = \frac{35 - 29}{0.3} = 20\ \mathrm{min}.$$

The standby chiller needs 15 minutes to start and load. The limit is reached before the standby unit carries load if $t_{\mathrm{limit}} < 15$ min, that is if $\dot{T} > 6/15 = 0.4\,\mathrm{K/min}$. The rate is itself uncertain, because it depends on the inferred airflow and on the thermal capacitance of the room and the chilled-water loop; with an illustrative $\dot{T} \sim \mathcal{N}(0.3, 0.06^2)\,\mathrm{K/min}$,

$$\Pr(T > T_{\max} \mid c_3) = \Pr(\dot{T} > 0.4) = 1 - \Phi\!\left(\frac{0.4 - 0.3}{0.06}\right) = 1 - \Phi(1.67) = 0.048.$$

A deterministic analysis says the facility has five minutes to spare. The probabilistic one says that in about one case in twenty the five minutes are not there.

24.3.4 Effective redundancy over time

Redundancy is a design label; the redundancy actually available changes with conditions. A plant designed as N+1 can be effectively N on a given afternoon: one chiller degraded, an extreme outdoor temperature derating the others, and another unit out for maintenance. The proposed use case recomputes the effective contingency tolerance as conditions evolve, instead of reading it from the design documents.

Figure 24.4 draws an illustrative plant of five chillers carrying a heat load of 7.2 MW. Each bar is the plant on one day, built up from the capacities of its available units; the vertical axis is cooling capacity in megawatts and the dashed red line is the load. The light top segment of each bar is the unit assumed lost in the contingency, the largest unit available. On the design

day each unit delivers 2.0 MW: four units carry the load and the fifth is the spare, so the loss of any one leaves 8.0 MW against 7.2 MW. On a hot day the condensers reject heat into hotter air and each healthy unit delivers 1.9 MW, while a degraded unit delivers 1.6 MW; the plant still survives a loss, with $9.2 - 1.9 = 7.3$ MW left, a margin of 0.1 MW. With one more unit in maintenance (dashed outline), the plant carries the load with 7.3 MW but the loss of any further unit leaves 5.4 MW, below the load. The light segment of the third bar straddles the load line: the plant is effectively N, with no contingency tolerance, although nothing in its design changed. The middle bar is the warning: the redundancy did not disappear in one step, it eroded to 0.1 MW first.

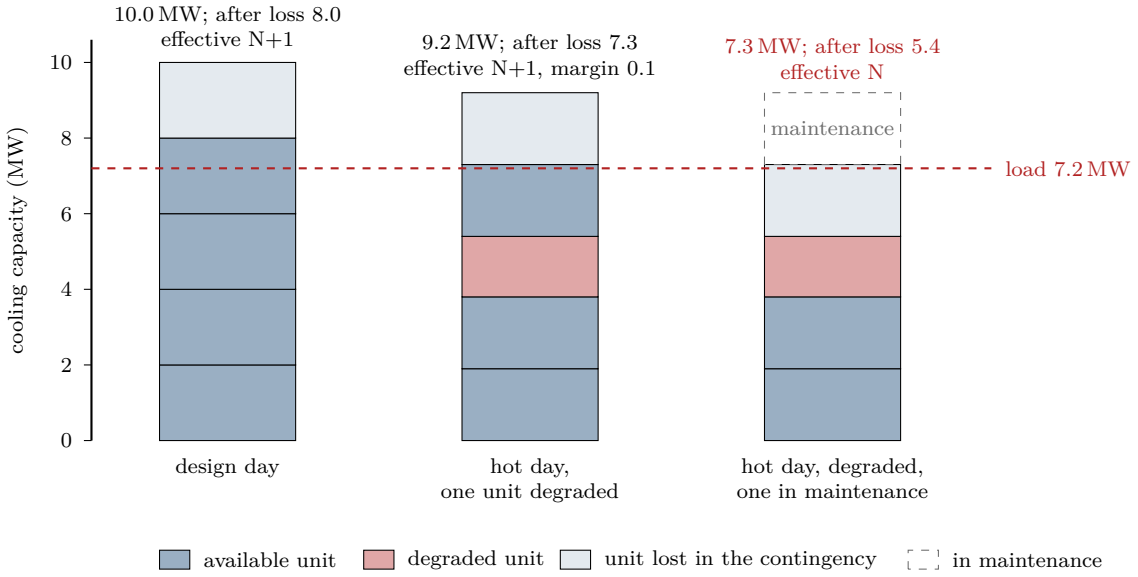


Figure 24.4: Effective redundancy of an illustrative five-chiller plant carrying 7.2 MW of heat. Design day: five units of 2.0 MW. Hot day: healthy units derated to 1.9 MW, one degraded unit at 1.6 MW. The light segment is the largest unit, assumed lost; where it straddles the load line, the plant has no contingency tolerance left.

24.3.5 Use case 7: risk-aware what-if

"What if 1 MW of IT load is added to hall B?" A deterministic what-if analysis (Chapter 18) runs the model once with nominal parameters and reports the hottest predicted inlet, $T_{\text{peak}} = 31.2$°C, below the 35°C limit. The proposed probabilistic analysis runs the same model over the posterior of its uncertain inputs and reports $\Pr(T_{\text{peak}} > 35\text{°C}) = 0.07$, together with the reason the uncertainty remains: the airflow in rows 8 to 10 is inferred, not measured. It adds the measurement that would narrow the answer: "if the airflow in rows 8 to 10 is verified, the estimate narrows substantially."

The two statements imply a spread. If T_{peak} is normal with mean 31.2°C, a tail probability of 0.07 above 35°C corresponds to $z = 1.476$ standard deviations, so

$$\sigma_T = \frac{35 - 31.2}{1.476} = 2.57\,\text{K}.$$

If verifying the airflow reduced the spread to an illustrative $\sigma_T = 1.5$ K with the same mean, the probability would fall to $1 - \Phi(3.8/1.5) = 1 - \Phi(2.53) = 0.006$, about one twelfth of its present value.

Figure 24.5 draws both distributions. The horizontal axis is the hottest inlet temperature after the change and the vertical axis its probability density. The solid curve is the present

distribution, with the airflow inferred; the dashed curve is the distribution after airflow verification. The vertical red line is the limit and the dotted line the deterministic prediction. The shaded area to the right of the limit under the solid curve is the 0.07 probability of violation; under the dashed curve the corresponding area is too small to see. The deterministic number is the same in both cases. What changes is whether the plan is acceptable: at a tolerance $\epsilon = 0.05$ the present answer is no and the answer after verification is yes.

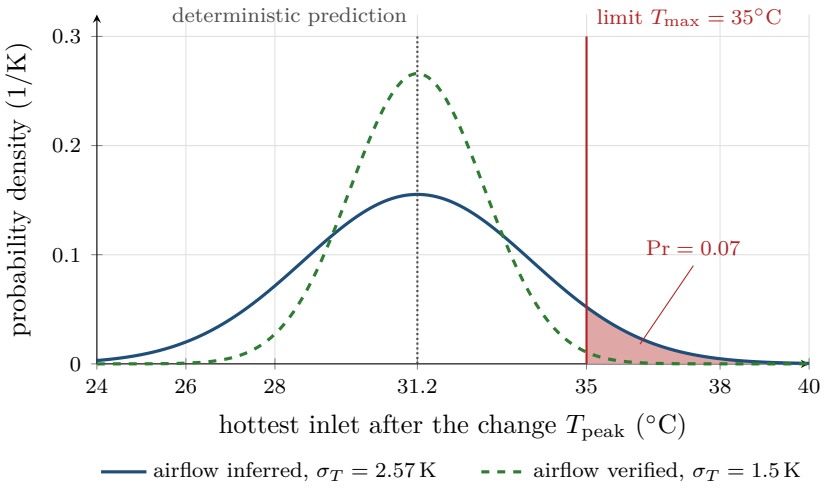


Figure 24.5: Risk-aware what-if for 1 MW added to hall B, illustrative. $T_{\text{peak}} \sim \mathcal{N}(31.2, \sigma_T^2)$ with $\sigma_T = 2.57\,\text{K}$ (airflow in rows 8 to 10 inferred) and $\sigma_T = 1.5\,\text{K}$ (airflow verified). The probability above $T_{\max} = 35°C$ is 0.07 and 0.006.

24.3.6 Use case 8: safe grid flexibility

A data center can offer flexibility to the grid by discharging its UPS batteries, deferring workload, capping server power or pre-cooling (Chapter 21). A statement such as "the battery can deliver 5 MW" is a nameplate statement. The proposed use case states instead the flexibility that can be delivered while the probability of compromising the facility stays low:

$$F_{95} = \max\ F \quad \text{subject to} \quad \Pr(\text{thermal, compute or backup violation} \mid F) < 0.05, \tag{24.5}$$

where a thermal violation is an inlet above its limit during pre-cooling or cooling curtailment, a compute violation is a missed workload deadline, and a backup violation is a battery state of charge below the reserve that the UPS must hold for a utility outage (Chapter 4). As an illustration, if the deliverable flexibility, after the backup reserve and the thermal and compute limits are respected, is $F \sim \mathcal{N}(4.4, 0.5^2)$ MW, then $F_{95} = 4.4 - 1.645 \times 0.5 = 3.6$ MW. The offer to the grid is 3.6 MW at 95 % confidence, not 5 MW. The structure is the same as that of the risk-qualified capacity of equation (24.3), with the flexibility in place of the load.

> **Principle 24.3.** *Every forward-looking quantity of a data center (additional capacity, contingency tolerance, the effect of a change, grid flexibility) is proposed here as a value at a stated probability of violation, with the sources of its uncertainty attributed and the measurement that would narrow it named.*

24.4 Boundaries of the approach

A framework is described as much by what it should not do as by what it should. This section lists the problems on which a physical-inference layer is not differentiated, states the boundary

in terms of what is known, draws it in two dimensions, and explains why the workflow and its integration matter more than any one method.

24.4.1 Problems where it is not differentiated

Five classes of problem are served well by existing tools, and the assessment of this chapter is that TERRA would add little to them:

- **Basic monitoring**: displaying temperatures, powers, battery state of charge and alarms. The BMS and DCIM do this.
- **Basic forecasting** of the IT power $P_{\mathrm{IT}}(t+h)$ at a horizon h. Standard time-series and machine-learning forecasters do this, and a physical model of the facility does not improve a forecast of the workload.
- **Pure CFD**: a detailed airflow field for a known geometry with known boundary conditions. A CFD solver does this, and a reduced-order network is less accurate by construction (Chapter 22).
- **Low-level controls**: PID loops, programmable logic controllers, chiller sequencing and safety interlocks (Chapter 10). These must be fast, deterministic and certified, and an inference layer belongs above them, not in them.
- **Basic optimization under known constraints**: choosing setpoints or a schedule when the model and its parameters are trusted. A standard solver does this.

Each of these lacks at least one of the four properties of Section 24.1.4. Monitoring and forecasting lack a physical model of the cause; CFD with known boundary conditions and optimization with known constraints lack the uncertainty; low-level control lacks competing explanations, since its job is to act on a measured error, not to explain it.

24.4.2 Known and uncertain problems

The boundary can be stated in terms of what is known. When the state is known, the model is known and the constraints are known, standard simulation or optimization suffices: run the model, or solve the problem. When the state is uncertain, the parameters are uncertain, the explanation of a fault is uncertain, and a decision is still required, a physical-inference layer is valuable, because no existing tool produces a decision-ready answer from that combination.

The test is whether uncertainty changes the answer. In the what-if example of Section 24.3.5, the deterministic prediction of 31.2°C said the plan is acceptable, and the probability of 0.07 said it is not, at a tolerance of 0.05. The uncertainty changed the decision, and that is the case in which the inference is worth its cost. Had the probability been 0.001, both analyses would have approved the plan and the inference would have added a number but not a decision.

> **Principle 24.4.** *Probabilistic inference matters when the uncertainty changes the answer. Where the deterministic and the probabilistic analysis lead to the same action, the simpler analysis suffices.*

24.4.3 Physical complexity and epistemic uncertainty

Figure 24.6 places problems on two axes. The horizontal axis is physical complexity: how many coupled mechanisms and spatial details the answer depends on. The vertical axis is epistemic uncertainty: how much of what the answer depends on is unknown. Each quadrant carries the kind of tool suited to it and two examples. Low complexity and low uncertainty is the

domain of standard analytics. Low complexity with high uncertainty calls for basic probabilistic methods, such as interval forecasts or failure statistics. High complexity with low uncertainty is the domain of simulation and optimization. The upper right quadrant, complex physics with incomplete knowledge, is where the four properties occur together, and it is the region proposed for TERRA.

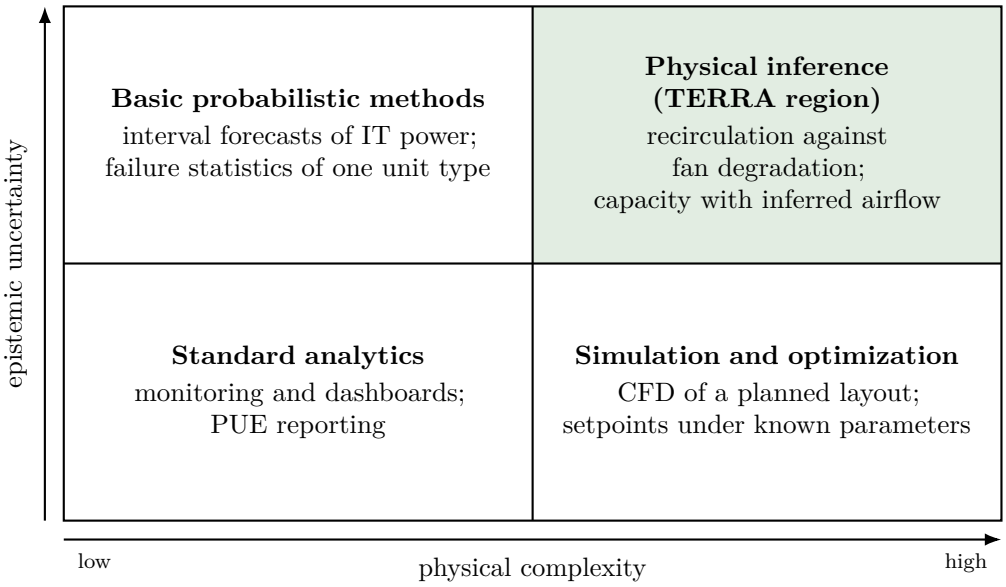


Figure 24.6: Problems placed by physical complexity and epistemic uncertainty. Each quadrant names the kind of tool suited to it and two examples. The shaded quadrant, complex physics with incomplete knowledge, is the region proposed for a physical-inference framework.

The figure also shows why the boundary moves. Instrumentation lowers epistemic uncertainty: a problem in the upper right moves down into the simulation quadrant once the airflow is measured. A reduced-order model lowers effective complexity: a problem in the lower right can sometimes be moved left. A physical-inference layer is therefore not permanently attached to a problem; it is attached to a problem in a given state of knowledge.

24.4.4 A workflow, not a single method

Computing the joint posterior $p(x, \theta, H \mid y)$ over states, parameters and hypotheses is one step. On its own it is ordinary Bayesian inference, and many tools perform it. The proposed workflow around it has five steps:

$$\text{physical model} \rightarrow \text{credibility} \rightarrow \text{inference} \rightarrow \text{risk} \rightarrow \text{next measurement or action.}$$

The physical model supplies the structure, the conservation laws and equipment relations of Chapters 1 to 8. The credibility step checks that the model is adequate and its parameters identifiable before anything is concluded from it (Section 24.2.5). The inference step computes the posterior. The risk step propagates the posterior to the probability of a violation (Section 24.3). The final step chooses the measurement or action with the largest value given that risk (Section 24.2.4).

A Bayesian posterior computed on a model that fails the credibility step is precise and wrong; a risk computed without a posterior is a guess with a number on it; a recommendation computed without a risk has no consequence to weigh. The packaged workflow matters more than any single method in it, because each step is only as useful as the one before it.

24.4.5 Integration around one question

Tools exist for every ingredient of the workflow: Bayesian inference libraries, fault-detection systems, CFD solvers, state estimators, optimizers and digital-twin platforms. The assessment of this chapter is that none of these components is new and that the distinctive feature of the proposal is their integration around one question:

what can actually be inferred from this physical system and these measurements, and what should be done with that knowledge?

The first half of the question is the identifiability analysis of Chapter 15 and the capability map of Chapter 23; the second half is the risk and decision analysis of Chapters 17 to 19. A facility that has an estimator but no credibility check, or a CFD model but no posterior over its boundary conditions, has the parts without the question that organizes them.

24.5 Outputs for operators

The value of an inference reaches an operator only through its outputs. This section describes the form proposed for them: conclusions classified by how they were obtained, evidence that can be traced, two compact cards for a diagnosis and a capacity, careful wording, and restraint stated as instrumentation guidance.

24.5.1 Classified conclusions

The distinction this book keeps between measured, calculated and inferred quantities becomes, in an output, a label on every conclusion. Table 24.3 lists five kinds of conclusion with an illustrative example of each for one rack, and the basis on which each rests. The first two are the book's measured and calculated quantities, here called observed and derived; the third is an inferred quantity; the fourth is an inferred cause, reported as a probability; the fifth is the explicit statement that the data do not support a conclusion.

Table 24.3: Five kinds of conclusion, with an illustrative example for one rack and the basis of each.

Kind	Example	Basis
Directly observed	$P_{\text{rack}} = 82\,\text{kW}$	a meter reading (measured)
Derived	$\dot{Q}_{\text{rack}} \approx 82\,\text{kW}$	the rack's electrical power becomes heat (calculated by a known relation)
Inferred	$\dot{m} = 3.1 \pm 0.4\,\text{kg/s}$	posterior mean and standard deviation of an unmeasured flow
Diagnostic hypothesis	$\Pr(\text{recirculation} \mid y) = 0.68$	posterior over competing causes
Unsupported	fan degradation against obstruction	the telemetry cannot distinguish them

The labels make the reliability of each line visible. The derived heat and the inferred airflow can also be checked against each other: by $\dot{Q} = \dot{m}\, c_{p,\text{air}}\, \Delta T$ they imply a temperature rise across the rack of $82/(3.1 \times 1.0) = 26.5\,\text{K}$, and between 23.4 and 30.4 K at one standard deviation of the airflow. If the rack's inlet and exhaust are both measured, a measured rise outside that range would contradict the inferred airflow and send the conclusion back to the credibility step. The last line of the table is a conclusion in its own right: it tells the operator that a question exists that the present telemetry cannot answer.

24.5.2 Evidence traceability

A probability of 0.68 for recirculation should not be accepted on authority. The proposed output lets the operator ask why and see the evidence. Table 24.4 lists, for this illustrative diagnosis, each piece of evidence, what was observed, and how it bears on the hypotheses of Section 24.2.2. The rows are taken from the signatures of Table 24.2: an unchanged rack power argues against higher IT load, an unchanged supply temperature against warmer chilled water, a raised neighboring hot-aisle temperature for recirculation, and pressure readings consistent with normal fan operation against fan failure as the cause of reduced airflow.

Table 24.4: Evidence behind an illustrative diagnosis $\Pr(H_{\text{recirc}} \mid y) = 0.68$.

Evidence	Observation	Bearing on the hypotheses
Rack power	unchanged	against higher IT load (H_1)
Supply air temperature	unchanged	against warmer chilled water (H_4)
Local inlet temperature	increased	the symptom; consistent with every hypothesis
Neighboring hot-aisle temperature	increased	for recirculation (H_3); not expected under a single-sensor bias (H_5)
Pressure across the fans	normal	against fan failure as the cause of reduced airflow (H_2)

Traceability serves two purposes. It lets an experienced operator disagree for a stated reason, for example by knowing that the neighboring sensor was recently moved, which is information the model did not have. And it exposes the evidence on which the conclusion rests most heavily, which is the evidence whose failure would change it.

24.5.3 The diagnostic card

The outputs of a diagnosis can be condensed into a card that an operator reads in a few seconds. Figure 24.7 draws the proposed card for an illustrative episode on rack 18, separate from the sensor-fault episode of Section 24.2.1. Each row has a label on the left, the content in the middle and, on the right, the kind of conclusion in the sense of Table 24.3. The first row is the observation, an inlet at 31.4°C. The second is the posterior over explanations, drawn as bars: hot-air recirculation at 68 %, reduced airflow at 22 %, sensor bias at 8 % and other causes at 2 %, which sum to one. The third row is the evidence; the fourth the risk, the probability that the inlet exceeds 35°C within 20 minutes, 0.11; the last the recommended next measurement, a direct measurement of the rack airflow, which is the one that would separate the first two explanations.

The card answers in order the questions an operator asks: what was seen, what probably caused it and what else could have, why, how serious it is, and what to do next. It does not state a single cause as fact, and it does not leave the operator with a probability and no action.

24.5.4 The capacity card and its wording

A capacity statement can be condensed in the same way. Figure 24.8 draws the proposed card for an illustrative facility, a different one from that of Section 24.3.2. Its rows are the additional IT capacity at 95 % confidence, 4.8 MW; the primary constraint, the airflow of hall C; the additional capacity that survives the loss of one unit, 2.1 MW; the dominant uncertainty, the distribution of rack airflow; and a what-if, that additional airflow verification could raise the supported capacity by 0.5 to 1.2 MW. The scale at the bottom of the card draws the three

Diagnostic card: rack 18 (illustrative)

Observed	rack 18 inlet temperature 31.4°C	*observed*
Explanations	hot-air recirculation 68 % reduced airflow 22 % sensor bias 8 % other 2 %	*hypothesis*
Evidence	normal rack power; normal CRAH supply temperature; elevated local exhaust temperature	*observed*
Risk	$\Pr(T > 35°\text{C within 20 min}) = 0.11$	*inferred*
Next measurement	direct rack airflow (separates recirculation from reduced airflow)	*recommendation*

Figure 24.7: A proposed diagnostic card for one episode on rack 18, illustrative. Rows: observation, posterior over explanations (bars, summing to 100 %), evidence, risk and the recommended next measurement. The right column states the kind of each conclusion.

capacities on one axis in megawatts: the N−1 capacity, the 95 % capacity and the range it could reach after verification, 5.3 to 6.0 MW.

Wording. The 4.8 MW of the card should be called a "model-supported capacity" or a "risk-qualified capacity", not a "certified capacity", unless a formal certification framework exists and has been applied. The phrase means a capacity whose uncertainty and model credibility have been explicitly evaluated: the probability is stated, the model has passed the credibility checks of Section 24.2.5, and the dominant source of uncertainty is named. "Certified" implies a guarantee and an external authority that the analysis does not provide.

24.5.5 Restraint as instrumentation guidance

The outputs of Sections 24.2.3 and 24.5.1 include conclusions that the data do not support. The proposed form of such a conclusion is not "cannot determine" but "not currently distinguishable", followed by "here is what would make it distinguishable". The first phrasing is a dead end; the second turns a limitation of the data into instrumentation guidance, and the guidance is itself a deliverable.

An example. The question is "is chiller 2 losing efficiency?" The efficiency of a chiller is its coefficient of performance, $\text{COP} = \dot{Q}_{\text{evap}}/W_{\text{comp}}$, and it depends on the condenser-water temperature and flow as well as on the machine's condition (Chapter 7). If the condenser-water flow is unmeasured, a fall in the computed COP can be explained by a degrading chiller or by a reduced condenser-water flow, and the two cannot be separated. The proposed output reads:

> "Chiller efficiency cannot be isolated from condenser-water effects because condenser-water flow is unmeasured. Recommended measurement: condenser-water flow or the pump operating point."

The pump operating point, its speed and differential pressure read against its curve, gives the flow indirectly (Chapter 9).

Capacity card (illustrative)

Additional IT	4.8 MW at 95 % confidence	*risk-qualified*
Primary constraint	airflow in hall C	*binding limit*
N−1 capacity	2.1 MW	*contingency*
Dominant uncertainty	rack airflow distribution	*attribution*
What-if	airflow verification could raise the supported capacity by 0.5 to 1.2 MW	*what-if*

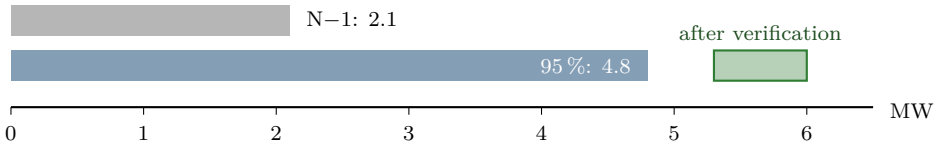


Figure 24.8: A proposed capacity card, illustrative. Rows: risk-qualified additional capacity, binding limit, contingency capacity, attribution of the uncertainty and a what-if. Scale: N−1 capacity 2.1 MW (gray), 95 % capacity 4.8 MW (blue) and the range 5.3 to 6.0 MW it could reach if the airflow were verified (green).

The value of the recommendation. For some clients the first question is not a diagnosis at all but: what instrumentation is needed before this question can be answered reliably? An analysis that answers it has converted uncertainty into an engineering plan, a list of measurements with the question each one makes answerable. The assessment of this chapter is that this is a deliverable with value of its own, available before any live inference is run, and it is the basis of the first phase of the adoption path that follows.

24.6 An adoption path

A framework of this kind would not be introduced into a facility all at once. This section sets out the path assessed as plausible: five phases, a first deliverable that needs no deployed software, one narrow live use case, a ladder of capabilities ordered by the model confidence they require, and a statement of what the framework is and is not.

24.6.1 Five phases

Table 24.5 lists the five phases with the question each answers and its output. Phase 1, the instrumentation audit, establishes what is measured, what is not and which questions are answerable. Phase 2 builds the smallest physical model adequate to those questions. Phase 3 tests that model's credibility, its adequacy and the identifiability of its parameters. Phase 4 runs operational inference: diagnosis, capacity, risk and what-if. Phase 5 repeats the risk and diagnosis calculations continuously as the facility operates.

The order matters. Each phase uses the result of the one before: the model of phase 2 is sized to the questions of phase 1; the credibility of phase 3 decides which of the inferences of phase 4 may be reported; and phase 5 is phase 4 repeated, with the credibility checks rerun as the facility and its instrumentation change.

Table 24.5: The five phases of the proposed adoption path.

Phase	Question	Output
1 Instrumentation audit	what is measured, what is not, which questions are answerable	measurement map, capability map, instrumentation priorities
2 Physical model	what is the smallest model adequate to the questions	coupled electrical, air and water network with its parameters
3 Credibility	is the model adequate; are its parameters identifiable	balance closure, identifiability analysis, validated parameter ranges
4 Operational inference	what is the state, the cause, the risk, the effect of a change	diagnoses, risk-qualified capacity, contingency risk, what-if
5 Continuous monitoring	how do state and risk evolve	repeated risk and diagnosis as data arrive

24.6.2 The instrumentation audit as a first deliverable

The first phase can be delivered without deploying any software in the facility. The proposed deliverable is a data center physical-inference and instrumentation audit with seven parts:

1. a system model: the electrical, air and water networks, their components and their topology;
2. a measurement map: which quantities are measured, where, at what rate and with what accuracy (Chapter 9);
3. a balance-closure analysis: which conservation balances can be checked from the telemetry and how well they close;
4. an inference capability map: which questions the present measurements can answer, as in Chapter 23;
5. the identifiable and non-identifiable quantities, with the reason each non-identifiable one fails (Chapter 15);
6. prioritized instrumentation recommendations, each tied to the question it would make answerable and ranked by value (Chapter 19);
7. candidate use cases, chosen from the eight of this chapter by what the audit shows to be answerable.

The audit is assessed as a sound first deliverable for three reasons: it is useful whatever the outcome, since a negative result ("this question cannot be answered with present instrumentation") is a finding; it requires only drawings, point lists and historical data; and it tests the framework's central claim, that physics and measurements together determine what can be known, before any live system depends on it.

24.6.3 One narrow live use case

After the audit, the proposed next step is one narrow live use case rather than a platform. Table 24.6 lists four candidates with what each needs and how its value would be shown. The assessment of this chapter is that risk-qualified cooling capacity (B) and airflow and thermal anomaly diagnosis (D) are the most compelling, for different reasons.

Table 24.6: Four candidates for a first live use case.

Use case	Needs	Value shown by
A Sensor trust	redundant evidence around each sensor	false alarms avoided; faulty sensors found
B Risk-qualified cooling capacity	credible thermal and airflow model; load forecast	usable megawatts gained or confirmed, at a stated probability
C Chiller-loss contingency risk	plant model; transient room model	effective redundancy and time to limit under today's conditions
D Airflow and thermal anomaly diagnosis	hypothesis set; physical signatures; faults to test against	correct cause ranked first on controlled or synthetic faults

Capacity is the commercially stronger case. A capacity result is stated in megawatts, and megawatts have a price. Two illustrative statements show the form: "the facility has 800 kW more safe IT capacity than the current conservative rule allows", and "airflow uncertainty prevents confident use of 1.2 MW of installed electrical capacity". The first is a direct business case for using capacity already paid for; the second is a direct business case for the instrumentation or remediation that would remove the uncertainty, since its cost can be set against the value of 1.2 MW of capacity.

Diagnosis is easier to demonstrate. A diagnosis can be tested before its financial impact is proven. A benchmark with controlled or synthetic faults, for example a sensor bias injected in software against a real airflow degradation created by partly blocking a filter, shows whether the correct hypothesis is ranked first and how often. The score is objective, the test can be repeated, and it requires no agreement about what a megawatt is worth. The assessment is that diagnosis is the natural research entry point and capacity the natural commercial one.

24.6.4 The maturity ladder

The capabilities of this chapter are not equally demanding. Figure 24.9 orders them as a ladder of six levels. Each box is one level, from measurement credibility at the bottom left through state estimation, probabilistic diagnosis, risk-qualified capacity and what-if decision support to supervisory optimization at the top right. The arrow beneath the ladder is the model confidence each level requires, which increases from left to right. Each level uses the outputs of the levels below it: a state cannot be estimated from sensors that are not credible, a fault cannot be diagnosed without an estimate of the normal state, a capacity cannot be risk-qualified without a posterior, and a change cannot be evaluated, or an action optimized, without all of them.

Optimization last. The ladder places optimization at the top for a reason. If the rack airflow is poorly identified, optimizing fan speed on a precise-looking airflow estimate is dangerous. Suppose the posterior is $\dot{m} = 3.1 \pm 1.0\,\mathrm{kg/s}$ for a rack generating 82 kW. The temperature rise across the rack, $\dot{Q}_{\mathrm{rack}}/(\dot{m}\, c_{p,\mathrm{air}})$, is 26.5 K at the posterior mean, but $82/4.1 = 20.0\,\mathrm{K}$ and $82/2.1 = 39.0\,\mathrm{K}$ one standard deviation either side. An optimizer that lowers fan speed until the nominal rise reaches a limit is choosing among outcomes that differ by a factor of two, and it will push hardest exactly where the estimate is weakest, since that is where the nominal model shows the most slack. The order is credibility, then inference, then risk, then optimization (Chapter 20).

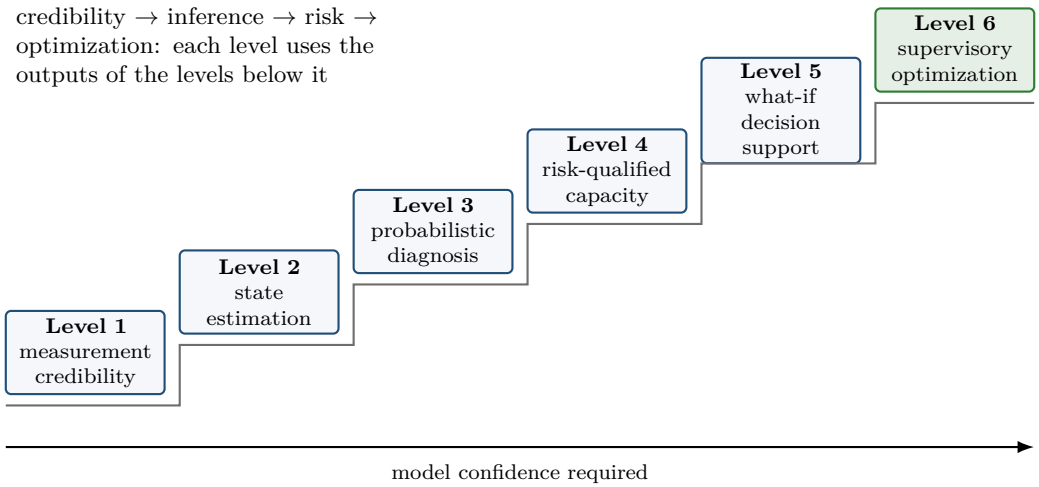


Figure 24.9: The proposed maturity ladder. Six levels of capability, each requiring the outputs of the levels below it and more confidence in the physical model than the level before.

The underlying position. Many systems begin with the question "what action optimizes performance?" The proposed framework begins with "is enough known about the physical system to justify that action?" The second question does not replace the first; it is asked before it, and its answer decides how high on the ladder the first may be asked.

> **Principle 24.5.** *Optimization is the last capability, not the first. An action is optimized only on a model whose credibility has been tested and whose uncertainty has been carried into the risk of the action.*

24.6.5 Product statement and claims

The chapter's assessment can be condensed into a product statement, quoted here in full:

> "TERRA helps data-center operators make physical-risk decisions when telemetry is incomplete, uncertain or contradictory. It combines engineering models with probabilistic inference to estimate hidden states, rank competing fault explanations, quantify operational risk, and identify the next measurement or action that would reduce uncertainty most."

Each clause of the statement corresponds to material of this chapter: the hidden states to Section 24.2.5 and the classified conclusions, the ranked explanations to Sections 24.2.2 and 24.5.2, the operational risk to Section 24.3, and the next measurement to Section 24.2.4.

What not to claim. Three claims would overstate the case: that the framework replaces digital twins, that it solves data center operations, or that it provides better CFD. It does none of these. It uses digital twins and CFD as inputs, it addresses one layer of operation, and its reduced-order models are less detailed than CFD by design. The claim the assessment supports is narrower: that it makes uncertain physical models decision-useful.

The stack revisited. Figure 24.1 drew the place of the framework in the facility's software: BMS, DCIM and sensors at the bottom; physical models, CFD and equipment models above them; the inference-and-risk layer, with its functions of credibility, state estimation, diagnosis, uncertainty, risk, next measurement and what-if, above those; and the operator, the optimizer

and the control system at the top. The adoption path of this section climbs that layer one function at a time. The framework is proposed as the inference-and-risk layer, not as the whole stack.

24.7 What to retain

Definition 24.1 (A physical-inference layer for data centers)**.** A probabilistic physical-inference layer is distinctive in a data center where physics, uncertainty, competing explanations and operational consequence occur together. Its proposed outputs are classified and traceable conclusions, risk-qualified quantities with attributed uncertainty, and the next measurement or action; it is introduced through an instrumentation audit and climbs a maturity ladder in which optimization comes last.

The relations to keep are the risk-qualified capacity $C_\epsilon = \max L$ subject to $\Pr(\text{violation} \mid L) < \epsilon$, its flexibility counterpart F_{95}, the attributed capacity $C_{95}^{(-k)} = \mu_C - z_{0.95}\,\sigma_C\sqrt{1 - s_k}$, the contingency risk $\Pr(T > T_{\max} \mid c)$ with its time to limit, and the flat-likelihood rule that returns the prior when the evidence cannot separate hypotheses. Table 24.7 lists the chapter's objects and the subsection that defines each.

24.8 Exercises

Exercise 24.1. A short telemetry report for one rack contains six statements: (i) "the PDU branch reads 41 kW"; (ii) "the rack rejects about 41 kW of heat"; (iii) "the inlet sensor reads 27.5°C"; (iv) "the rack airflow is 2.0 ± 0.3 kg/s"; (v) "the probability that the rack's inlet sensor is biased is 0.15"; (vi) "the rise in exhaust temperature is caused by a clogged filter, not by a slower fan", where neither fan speed, fan pressure nor filter pressure drop is measured. (a) Classify each statement as observed, derived, inferred, diagnostic hypothesis or unsupported, using Table 24.3, and state the basis of each. (b) Using $c_{p,\text{air}} = 1.0\,\text{kJ/(kg K)}$, compute the temperature rise across the rack implied by (ii) and (iv), and its range at one standard deviation of the airflow. (c) Rewrite (vi) in the form recommended in Section 24.5.5, naming a measurement that would make the two causes distinguishable.

Exercise 24.2. A risk-qualified capacity is $C_{95} = 3.4$ MW with $\sigma_C = 0.5$ MW, and its variance is attributed to airflow (50 %), the plant state (25 %), the load forecast (15 %) and sensors (10 %), independently. (a) Compute μ_C with $z_{0.95} = 1.645$. (b) Compute C_{95} if each source in turn is removed, using equation (24.4), and rank the gains. (c) Compute C_{95} if the airflow variance is halved and the sensor variance is removed at the same time. (d) Compute C_{99} before any improvement, with $z_{0.99} = 2.326$, and explain why it is lower than C_{95}.

Exercise 24.3. Place each of the following problems in one quadrant of Figure 24.6, say which kind of tool suits it, and state which of the four properties of Section 24.1.4 it lacks, if any: (a) computing the monthly PUE of a facility from its utility and UPS output meters; (b) the airflow field of a new hall whose geometry, tile layout and air-handler settings are all specified; (c) deciding whether a rising chiller compressor power is caused by condenser fouling or by low condenser-water flow, when the flow is unmeasured and a planned load increase depends on the answer; (d) estimating the probability that at least one of 400 identical server power supplies fails in the next month, from the vendor's failure rate. For (c), name the measurement that would move the problem to a different quadrant, and say which one.

Exercise 24.4. A chiller plant has four units of 2.5 MW each at design conditions and carries 6.8 MW of heat. (a) State its design redundancy level. (b) On a hot day each healthy unit delivers 2.3 MW and one degraded unit delivers 1.9 MW; compute the available capacity, the

Table 24.7: The objects of this chapter and where each is defined.

Object	Expression or content	Role	Defined in
Tool landscape	Table 24.1	what exists already	24.1.1
Inference layer	Figure 24.1	between telemetry and decisions	24.1.3
Four properties	physics, uncertainty, competing explanations, consequence	where the layer is distinctive	24.1.4
Sensor trust	equation (24.1)	credibility of a reading	24.2.1
Flat likelihood	equation (24.2)	restraint when data cannot separate	24.2.3
Next-best measurement	IG over a tied posterior	breaks an ambiguity	24.2.4
Model credibility	six questions	precondition for every inference	24.2.5
Risk-qualified capacity	C_ϵ, equation (24.3)	usable load at a stated risk	24.3.1
Attributed uncertainty	equation (24.4)	which knowledge to improve first	24.3.2
Contingency risk	$\Pr(T > T_{\max} \mid c)$, t_{limit}	today's consequence of a failure	24.3.3
Effective redundancy	Figure 24.4	redundancy under today's conditions	24.3.4
Risk-aware what-if	$\Pr(T_{\text{peak}} > T_{\max})$	a change with its risk	24.3.5
Safe flexibility	F_{95}, equation (24.5)	grid offer at a stated risk	24.3.6
Complexity–uncertainty map	Figure 24.6	where each tool belongs	24.4.3
Classified conclusions	Table 24.3	label on every output	24.5.1
Diagnostic and capacity cards	Figures 24.7, 24.8	compact operator outputs	24.5.3
Adoption phases	Table 24.5	audit first, monitoring last	24.6.1
Maturity ladder	Figure 24.9	optimization last	24.6.4

capacity after the loss of the largest unit, and the effective redundancy. (c) The same day one healthy unit goes into maintenance; repeat (b). (d) In case (b), the loss of one unit would make the hottest inlet, now at 28°C, rise at $\dot{T} \sim \mathcal{N}(0.25, 0.05^2)$ K/min toward $T_{\max} = 35$°C, and the standby unit needs 20 minutes to carry load. Compute the nominal t_{limit} and $\Pr(T > T_{\max} \mid c)$ before the standby unit carries load.

24.9 The book in one page

The eight use cases and the six levels of the maturity ladder each rest on material developed earlier in the book. Table 24.8 maps each of them to the chapters that develop it. The use cases are listed in the order in which they were introduced; the levels in the order of the ladder. Every chapter of the book appears at least once, because each supplies a piece that one of the use cases needs: the physics of the facility, its measurement, its failure and redundancy, or the methods of inference and decision built on them.

Table 24.8: The eight use cases and the six levels of the maturity ladder, with the chapters that develop each.

Item	What it requires	Chapters
Use cases		
1 Sensor trust	redundant physical evidence; bias against noise; thermal time constants	8, 9, 13, 16
2 Competing cooling diagnoses	airflow and cooling physics; signatures; posterior over hypotheses	6, 7, 13, 15, 16
3 Risk-qualified capacity	coupled limits; chance constraints; attribution	3, 5, 14, 17
4 Contingency risk	redundancy; backup power; transient room response	4, 8, 12, 18
5 Next-best measurement	entropy, information gain, value of information	9, 19, 23
6 Model credibility	balance closure; identifiability; model fidelity	5, 11, 15, 22
7 Risk-aware what-if	scenario propagation under uncertainty	17, 18
8 Safe grid flexibility	battery reserve; workload deadlines; pre-cooling	2, 4, 20, 21
Maturity ladder		
Level 1 Measurement credibility	sensors, balances, available data	1, 9, 11, 23
Level 2 State estimation	estimators; identifiability	14, 15
Level 3 Probabilistic diagnosis	failure modes; posterior over causes	13, 16
Level 4 Risk-qualified capacity	reliability; capacity under risk	12, 17
Level 5 What-if decision support	scenarios; measurement choice; model fidelity	18, 19, 22
Level 6 Supervisory optimization	controls; optimization; grid interaction	10, 20, 21

The conclusion of the book follows from that table. The physics of a data center is not exotic: electrical power becomes heat in the IT equipment, the heat is carried by air and water, and it is rejected to the environment (Chapters 1 to 8). The hard operational problem is that the facility rarely observes those processes completely. Its sensors are sparse and some are biased; its airflows, recirculation fractions and coil conductances are hidden states; its power topology is uncertain; its equipment degrades; and its weather, load and maintenance state change from hour to hour (Chapters 9 to 15). The operator must nonetheless answer the same questions every day: is something wrong; what is it; how serious is it; how much capacity remains; what happens if another component fails; what should be measured next; and what should be done (Chapters 16 to 21). Data centers are therefore a plausible application area for a framework of this kind, not because they need another simulation tool, but because they need trustworthy inference and physical-risk decisions under incomplete information.