

Probabilistic Modeling of Data Centers

Cooling, capacity and diagnosis

390 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Probabilistic Modeling of Data Centers

Cooling, Capacity and Diagnosis

Ricardo Marquez

Working draft

Preface

This book is written for an engineer who knows conservation laws and linear algebra and has never met a probabilistic graphical model. It takes one idea and follows it through: that the structure of a physical system, the way conserved quantities accumulate at places and move between them and cross the system's boundary, is also the structure of every probabilistic question one can ask about that system. A model of the physics is a graph. A model of what we do not know about the physics is the same graph with probability attached to it. Inference on that graph is how a measurement in one place becomes knowledge about another, how a failure is diagnosed, how a decision is justified.

This edition takes its examples from the data center: the chassis and the rack, the row and its containment, the hall and its air, the chilled-water plant, and the campus that pays the bill. A data center is a good system to point that idea at, for a reason the book makes precise. It is a hierarchy that was drawn by engineers — device inside chassis inside rack inside row inside hall inside building inside campus — and the levels nest because the cuts nest. It conserves several quantities at once on the same graph: heat, air, water and electrical power. Its instruments are sparse relative to what an operator wants to know, so most of the questions are inference questions rather than reading questions. And the decisions it forces — where to put the next server, whether the economizer pays, which of two identical racks carries the hidden load — are decisions under uncertainty with money and uptime attached.

Every system worth modeling is a piece cut out of a larger one, trading a conserved quantity with an environment it does not model. For a data center the pieces nest: a server trades air and power with its rack, a rack with its row, a contained aisle with its hall, and the hall with the chilled-water plant and the utility feed. At each level the boundary is where one model stops and the next begins, and it is where the modeler's knowledge is thinnest. As the later chapters show, it is also where the mathematics of inference does its most useful work, and the nesting of the boundaries is why the hierarchy is also a good way to divide the computation. Chapter 1 names such a system a *boundary-exchanging system*.

Nothing in the method of Parts I to V is specific to data centers. The balances, closures and ports are the same whether the conserved quantity is heat, air, water or power, and a data center conserves all four on one graph; Table 1.3 lines them up. Three companion editions take the same method text through electric power transmission, water distribution, and district heating and cooling. Each is complete in itself, and a reader of this one needs none of them.

The Introduction says why the book exists, what it inherits from four older traditions, what it adds, and how to read it. Part I is deterministic and establishes the objects. Part II attaches probability to them and states, rather than skips, the places where that attachment needs care. Part III is inference. Part IV asks the questions an operator actually asks. Part V handles time, hierarchy and scale. Part VI is ten complete studies at every level of the hierarchy, from which of two identical racks carries the load to how much of a hall's capacity can actually be used, and

each can be reproduced from the material in the earlier parts. Each chapter ends with a short section of notes that says which of its ideas are inherited and from where, and which are the book's own.

Contents

Preface **ii**

0 Introduction **1**
0.1 The questions a data center operator asks . . . 1
0.2 Where the graph comes from . . . 2
0.3 One subject, not four . . . 3
0.4 Probability in the structure, not around the solver . . . 5
0.5 The boundary . . . 6
0.6 From "what is the state" to "what can I know" . . . 6
0.7 What is inherited and what is the book's . . . 7
0.8 How to read this book . . . 7
0.9 What this book does not do . . . 9
Notes and further reading . . . 9

I Physical systems as graphs **10**

1 Conservation, closure, and the boundary **11**
1.1 Three questions about any system . . . 11
1.2 Nodes conserve . . . 12
1.2.1 The balance law . . . 12
1.2.2 Node roles . . . 12
1.2.3 Incidence . . . 13
1.3 Edges close . . . 13
1.3.1 Closure relations . . . 14
1.3.2 A closure is a relation, not an assignment . . . 14
1.3.3 Parameters . . . 15
1.4 The boundary . . . 15
1.4.1 Reservoirs and imposed conditions . . . 15
1.4.2 Exactness at a cut . . . 16
1.4.3 Why the boundary matters most . . . 17
1.5 The residual system . . . 17
1.5.1 Well-posedness . . . 18
1.5.2 Sparsity . . . 18
1.6 Three worked examples . . . 19
1.7 In practice . . . 26

1.8 What is missing . . . 27
Notes and further reading . . . 28
Summary . . . 28
Exercises . . . 29
Solutions to selected exercises . . . 29
Problems from the studies . . . 30

2 From equations to factor graphs **31**
2.1 The rows as a graph . . . 31
2.2 Three languages . . . 32
2.2.1 Directed graphs . . . 32
2.2.2 The directionality problem . . . 33
2.2.3 Undirected graphs . . . 35
2.2.4 Factor graphs . . . 35
2.3 The physical system as a factor graph . . . 36
2.3.1 A bus is not a variable . . . 37
2.3.2 Two domains on one graph . . . 37
2.4 The factor graph is not unique . . . 38
2.5 Vocabulary . . . 39
2.6 In practice . . . 40
2.7 What is missing . . . 41
Notes and further reading . . . 41
Summary . . . 42
Exercises . . . 43
Solutions to selected exercises . . . 43
Problems from the studies . . . 44

II Probability on a physical graph **45**

3 Where probability enters **46**
3.1 A probability on the state . . . 46
3.2 Five kinds of factor . . . 47
3.2.1 Priors . . . 47
3.2.2 Closure factors . . . 48
3.2.3 Conservation factors . . . 48
3.2.4 Measurement factors . . . 49
3.2.5 Failure factors . . . 49
3.2.6 The complete list . . . 50
3.3 The joint as an objective . . . 50
3.3.1 Residuals in units of their widths . . . 51
3.3.2 Conditioning a Gaussian on one reading . . . 51
3.4 A worked example: one sensor, then two . . . 54
3.5 A second example: the triangle with a flow meter . . . 56
3.6 Reading the graph . . . 59
3.7 In practice . . . 60
3.8 What is missing . . . 61

Notes and further reading . . . 61
Summary . . . 62
Exercises . . . 62
Solutions to selected exercises . . . 62
Problems from the studies . . . 63

4 Hard constraints, soft constraints, and the manifold 64
4.1 The hard joint and its manifold . . . 64
4.2 Conditioning on the manifold . . . 65
4.3 The soft joint . . . 66
4.4 Why soften . . . 67
4.5 Choosing the widths . . . 68
4.6 Worked example: the chain with soft physics . . . 70
4.7 A second example: the triangle with an unmodeled load . . . 72
4.8 In practice . . . 75
4.9 What is missing . . . 75
Notes and further reading . . . 76
Summary . . . 76
Exercises . . . 77
Solutions to selected exercises . . . 77
Problems from the studies . . . 78

5 Factorization, separators, and treewidth 79
5.1 Factorization is not independence . . . 79
5.2 Reading independence off the graph . . . 80
5.3 Explaining away, precisely . . . 81
5.4 Ports are separators . . . 82
5.5 Elimination and fill . . . 83
5.5.1 One elimination . . . 83
5.5.2 Order and width . . . 84
5.5.3 What this means for a physical graph . . . 85
5.6 Worked examples: counting width and fill . . . 86
5.7 Why separators matter for cost . . . 88
5.8 In practice . . . 90
5.9 What is missing . . . 91
Notes and further reading . . . 91
Summary . . . 91
Exercises . . . 92
Solutions to selected exercises . . . 92
Problems from the studies . . . 93

6 The linear-Gaussian case 94
6.1 The Gaussian in information form . . . 94
6.2 The most probable state . . . 95
6.2.1 Gauss–Newton against Newton . . . 96
6.3 Elimination is Cholesky . . . 96
6.3.1 What the factorization costs . . . 97

6.4 The Laplace posterior . . . 99
6.5 Worked example: the chain in matrix form . . . 99
6.6 Second worked example: the triangle in matrix form . . . 102
6.7 Filters, fields, and state estimators . . . 104
6.8 In practice . . . 105
6.9 What is missing . . . 106
Notes and further reading . . . 106
Summary . . . 106
Exercises . . . 107
Solutions to selected exercises . . . 107
Problems from the studies . . . 108

7 Beyond Gaussian: hybrid states and nonlinearity 110
7.1 Four ways out of the Gaussian world . . . 110
7.2 Hybrid states: the mixture over branches . . . 111
7.2.1 Worked example: a feeder that may be out . . . 112
7.2.2 Many discrete variables . . . 113
7.3 Piecewise closures and regimes . . . 114
7.3.1 Worked example: a thermostat . . . 114
7.4 Strong nonlinearity . . . 116
7.5 Bounded parameters . . . 117
7.6 What survives . . . 118
7.7 In practice . . . 119
7.8 What is missing . . . 119
Notes and further reading . . . 120
Summary . . . 120
Exercises . . . 121
Solutions to selected exercises . . . 121
Problems from the studies . . . 123

III Inference 124

8 Exact inference: elimination, junction trees, Schur complements 125
8.1 What exact means . . . 125
8.2 Messages on a tree . . . 125
8.3 Loops: the junction tree . . . 127
8.4 Regions and ports: the Schur complement as a message . . . 128
8.5 Reuse . . . 130
8.6 Worked example: two regions, one cross-tie . . . 131
8.7 Second worked example: two halls and a wall . . . 132
8.8 Discrete variables by region . . . 134
8.9 In practice . . . 135
8.10 What is missing . . . 136
Notes and further reading . . . 136
Summary . . . 137
Exercises . . . 138

Solutions to selected exercises . . . 138
Problems from the studies . . . 140

9 Approximate inference 141
9.1 When exact is too expensive . . . 141
9.2 Loopy belief propagation . . . 142
9.2.1 On the book's models . . . 142
9.3 Variational methods and expectation propagation . . . 144
9.4 Simulation inside a region . . . 146
9.4.1 What a mixture message loses when collapsed . . . 147
9.4.2 Worked example: a chilled-water plant as a simulated region . . . 147
9.5 Pruning by evidence . . . 149
9.6 The hierarchy, and how to choose . . . 150
9.7 In practice . . . 151
9.8 What is missing . . . 152
Notes and further reading . . . 152
Summary . . . 152
Exercises . . . 153
Solutions to selected exercises . . . 154
Problems from the studies . . . 155

10 Enumeration, Monte Carlo, and factorized inference 156
10.1 The question, stated honestly . . . 156
10.2 The problem . . . 157
10.3 Enumeration: the product grid . . . 157
10.4 Monte Carlo . . . 159
10.5 Attribution: where the structural difference appears . . . 160
10.6 The factorized route: cuts of the dependency graph . . . 161
10.7 Second worked example: two chilled-water plants . . . 162
10.8 Beyond three inputs . . . 166
10.9 The claim . . . 166
10.10In practice . . . 167
10.11What is missing . . . 168
Notes and further reading . . . 168
Summary . . . 168
Exercises . . . 169
Solutions to selected exercises . . . 169
Problems from the studies . . . 170

11 Sequential evidence 171
11.1 Readings arrive one at a time . . . 171
11.2 One reading, one rank-one update . . . 171
11.3 The innovation . . . 172
11.4 Many readings . . . 173
11.5 Worked example: a stream of flow readings . . . 174
11.6 The Laplace posterior and re-linearization . . . 175
11.7 The value of a reading before it is taken . . . 176

11.8 Cost at scale . . . 178
11.9 Worked example: where to put the next thermometer . . . 179
11.10In practice . . . 180
11.11What is missing . . . 181
Notes and further reading . . . 181
Summary . . . 182
Exercises . . . 182
Solutions to selected exercises . . . 183
Problems from the studies . . . 184

IV The operator's questions 185

12 Conditioning, diagnosis, and virtual sensors 186
12.1 Everything is a readout . . . 186
12.2 Virtual sensors . . . 186
12.3 Is the model adequate? . . . 187
12.4 Where is it wrong? . . . 188
12.5 Worked example: bad data on a power network . . . 189
12.6 Which explanation? . . . 190
12.7 Can this parameter be recovered? . . . 191
12.8 Worked example: can fouling be seen? . . . 192
12.9 Which sensor next? . . . 193
12.10Should the Gaussian be believed? . . . 195
12.11In practice . . . 197
12.12What is missing . . . 197
Notes and further reading . . . 197
Summary . . . 198
Exercises . . . 199
Solutions to selected exercises . . . 199
Problems from the studies . . . 200

13 Interventions and counterfactuals 201
13.1 Why the graph is not enough . . . 201
13.2 Rows are mechanisms; inputs are exogenous . . . 201
13.3 Seeing versus doing, on the chain . . . 202
13.4 Counterfactuals: three steps . . . 204
13.5 Decisions as branches . . . 207
13.6 Worked example: an afternoon heat wave . . . 208
13.7 Worked example: holding the tie flow . . . 210
13.8 What the semantics does not decide . . . 211
13.9 In practice . . . 211
13.10What is missing . . . 212
Notes and further reading . . . 212
Summary . . . 213
Exercises . . . 213
Solutions to selected exercises . . . 214

Problems from the studies . . . 215

14 Attribution, sensitivity, and certificates 216
14.1 Three questions, one derivative . . . 216
14.2 Sensitivities from the solve . . . 216
14.3 Attribution . . . 217
14.4 A census with gradients . . . 219
14.4.1 What the census answers . . . 220
14.4.2 The three-feeder problem . . . 220
14.5 The kink . . . 222
14.6 Worked example: convexity lost and recovered . . . 223
14.7 Conditions . . . 226
14.8 In practice . . . 227
14.9 What is missing . . . 227
Notes and further reading . . . 227
Summary . . . 228
Exercises . . . 229
Solutions to selected exercises . . . 229
Problems from the studies . . . 230

15 Combinatorial failure 231
15.1 The question . . . 231
15.2 The prior over failure sets . . . 232
15.2.1 Independent failures . . . 232
15.2.2 Failure as an intervention . . . 233
15.3 The consequence per branch, and how to get it cheaply . . . 233
15.4 Worked example: a six-bus distribution . . . 234
15.5 The posterior over failure sets . . . 236
15.6 Enumeration by regions . . . 237
15.7 The worst k-set as a query . . . 238
15.8 Common cause and correlated failure . . . 239
15.8.1 A shared shock . . . 239
15.8.2 A mixture over the rate . . . 239
15.9 Worked example: a cooled room . . . 240
15.10In practice . . . 242
15.11What is missing . . . 242
Notes and further reading . . . 243
Summary . . . 243
Exercises . . . 244
Solutions to selected exercises . . . 244
Problems from the studies . . . 245

V Time, hierarchy, and scale 246

16 Dynamics and temporal factors 247
16.1 The accumulation term made live . . . 247

16.2 Integration as a sequence of steady solves . . . 248
16.3 The unrolled graph . . . 249
16.4 Process noise . . . 250
16.5 Filtering, smoothing, and prediction as eliminations . . . 251
16.6 Worked example: tracking a drifting source . . . 252
16.7 Detecting a change in time . . . 253
16.8 Two hypotheses the steady state cannot separate . . . 254
16.9 Events and cascades . . . 255
16.10Worked example: the conductors of the six-bus distribution . . . 256
16.11In practice . . . 258
16.12What is missing . . . 259
Notes and further reading . . . 259
Summary . . . 259
Exercises . . . 260
Solutions to selected exercises . . . 261
Problems from the studies . . . 261

17 Hierarchy as physical decomposition 262
17.1 One conservation law used twice . . . 262
17.2 Nested messages . . . 263
17.3 What the parent needs to know . . . 264
17.4 Worked example: a machine room in three levels . . . 265
17.5 The cut, three times . . . 269
17.6 Worked example: a rack row under a distribution unit . . . 270
17.7 Hierarchy in time . . . 271
17.8 In practice . . . 273
17.9 What is missing . . . 274
Notes and further reading . . . 274
Summary . . . 274
Exercises . . . 275
Solutions to selected exercises . . . 275
Problems from the studies . . . 276

18 Partitioning at scale 277
18.1 The cost of a cut . . . 277
18.2 Elimination width of physical networks . . . 278
18.3 Choosing the cut . . . 279
18.4 Coupling the regions: messages, propagation, or iteration . . . 281
18.5 Enumeration by region . . . 284
18.6 The machinery is not this domain's . . . 287
18.7 In practice . . . 288
18.8 What is missing . . . 288
Notes and further reading . . . 288
Summary . . . 289
Exercises . . . 289
Solutions to selected exercises . . . 290
Problems from the studies . . . 291

VI Case studies **292**

19 Case studies **293**
19.1 What a case study is here . . . 293
19.2 The studies . . . 294
19.3 Reading a case study . . . 294

20 Three routes on two halls **295**
20.1 The question . . . 295
20.2 The system . . . 296
20.3 The three routes . . . 296
20.4 Results . . . 297
20.5 What surprised . . . 299
20.6 Reproducing the run . . . 299
Notes and further reading . . . 300

21 Which of two identical racks **301**
21.1 The question . . . 301
21.2 The system . . . 301
21.3 Two representations of the same posterior . . . 302
21.4 Results . . . 302
21.5 What surprised . . . 304
21.6 Reproducing the run . . . 305
Notes and further reading . . . 306

22 Where the Gaussian band stops **307**
22.1 The question . . . 307
22.2 The system . . . 307
22.3 Three things are computed at each point . . . 308
22.4 Results . . . 308
22.5 What surprised . . . 311
22.6 Reproducing the run . . . 311
Notes and further reading . . . 312

23 The expensive chiller nobody sees **313**
23.1 The question . . . 313
23.2 The plant . . . 313
23.3 Method . . . 314
23.4 Results . . . 314
23.5 What surprised . . . 315
23.6 Reproducing the run . . . 317
Notes and further reading . . . 317

24 When one explanation is not enough **318**
24.1 The question . . . 318
24.2 The hypotheses . . . 318
24.3 Results . . . 319
24.4 What surprised . . . 320

24.5 Reproducing the run . . . 321
Notes and further reading . . . 321

25 Telemetry you should not trust 322
25.1 The question . . . 322
25.2 Four corruptions . . . 322
25.3 Results . . . 323
25.4 What surprised . . . 324
25.5 Reproducing the run . . . 325
Notes and further reading . . . 325

26 Which sensor to install next 326
26.1 Which reading to take now . . . 326
26.2 Which four to install . . . 327
26.3 What surprised . . . 328
26.4 Reproducing the runs . . . 329
Notes and further reading . . . 330

27 How much capacity can we actually use 331
27.1 The question . . . 331
27.2 The building . . . 331
27.3 Results . . . 332
27.4 What surprised . . . 333
27.5 Reproducing the run . . . 334
Notes and further reading . . . 334

28 What the graph is worth 335
28.1 The question . . . 335
28.2 The system . . . 335
28.3 Results . . . 336
28.4 What surprised . . . 337
28.5 Reproducing the run . . . 338
Notes and further reading . . . 338

29 Same heatwave, four resolutions 339
29.1 The event and the facility . . . 339
29.2 What each resolution can see . . . 339
29.3 What each resolution does . . . 340
29.4 What surprised . . . 341
29.5 Reproducing the run . . . 342
Notes and further reading . . . 342

A Notation 343
A.1 Conventions . . . 343
A.2 Symbols . . . 344
A.3 Letters with more than one meaning . . . 347

B Gaussian identities **349**
B.1 The two forms of a Gaussian . . . 349
B.2 Block matrices . . . 350
B.3 Marginals and conditionals . . . 351
B.4 Products and linear-Gaussian models . . . 352
B.5 Rank-one changes . . . 353
B.6 Integrals, quadratic forms and divergences . . . 353
B.7 The Laplace approximation . . . 354
B.8 Where the book uses each identity . . . 355

C From the book to the engine **356**
C.1 What TERRA is . . . 356
C.2 A model in a dozen lines . . . 357
C.3 Where each construction lives . . . 357
C.4 The case studies . . . 358

D A catalogue of closures **360**
D.1 How to read an entry . . . 360
D.2 The four patterns in the engine . . . 362
D.3 Thermal . . . 362
D.4 Fluid . . . 364
D.5 Moist air . . . 366
D.6 Electrical . . . 366
D.7 Storage and electrochemistry . . . 368
D.8 Plant equipment . . . 368

Chapter 0

Introduction

An inlet-temperature alarm trips on one rack in the middle of a row. Its neighbors are cool, and the cooling units report normal supply air. The list of explanations is short. Hot exhaust is finding its way back to the rack's front, over the top of the row or through a gap where a blanking panel should be. Or a cooling coil has fouled and is delivering less than its setpoint implies. Or the inlet sensor has been moved, or has drifted. Or the new deployment in that rack is drawing more than its nameplate. The facility knows its electrical load at the power distribution units and its air temperatures at a few points per row, and the decision that follows has money and uptime attached: move the load, add airflow, or accept the risk.

The engineer on shift already has the physics of that hall: an energy and an air balance at every rack, aisle and plenum, a conductance or a fan curve on every path the air can take, a coil effectiveness at every cooling unit, and the chilled-water supply temperature imposed. Given the loads, a solver turns those equations into every temperature and every airflow. What the solver cannot do is run the other way: take a row's worth of sensor readings and say which of the explanations they support, how strongly, what the inlet temperature is at the racks nobody measures, and whether the row can take the next deployment. Probabilistic graphical models supply that machinery. They are usually taught apart from conservation laws, constitutive relations and engineering solvers, and the engineer who wants both has had to learn two subjects and build the bridge alone. This book is the bridge. Its claim is that the bridge is short, because the graph the probabilistic machinery needs is already there in the equations the engineer wrote, and that once the graph is recognized, every question on the shift's list is a readout of one object. This chapter says why that claim is worth a book, what it inherits from four older traditions, what it adds, and how to read it.

0.1 The questions a data center operator asks

A model of a physical system, as engineering teaches it, is a set of equations and a solver. The equations are the balances, the closures and the boundary conditions; the solver turns known inputs into a computed state. The pipeline is

$$\text{inputs} \longrightarrow \text{solver} \longrightarrow \text{outputs}, \tag{0.1}$$

and it answers one question well: given these inputs, what is the state? Most textbooks on physical modeling stop there, and for design that is often enough.

Operation asks different questions. A data center is instrumented for billing and for alarms more than for understanding: power at the distribution units and sometimes at the outlets, air temperatures at a few heights on a few racks per row, supply and return temperatures at each cooling unit. The sensors are noisy, some are misplaced, and none reads the airflow that decides where the heat goes. Coils foul. Fans wear. Loads arrive with a nameplate rating that is a ceiling, not a measurement. The questions an operator or a capacity planner actually asks are:

- What are the temperatures and airflows across the hall now, given a few sensors per row and a model I only partly trust?
- Has that cooling unit's coil fouled, and by how much?
- What is the inlet temperature at the racks nobody measures?
- Which of two identical racks carries the hidden load, and is the hot inlet recirculation, a fouled coil, a misplaced sensor or real load?
- How uncertain is each of those answers?
- Where should the next sensor go, and what would it be worth?
- What would happen if I moved that load, raised the supply-air setpoint, or took that cooling unit out for maintenance?
- Which uncertain load or airflow drives the risk of an inlet excursion, whether the economizer pays today, and is the number I am about to report a floor, a ceiling or a guess?

None of these is a question the pipeline (0.1) can answer. Each asks the model to run in a direction other than inputs to outputs, or to weigh a reading against a law, or to compare explanations, or to say how far its own answer can be trusted. The pipeline has no place to put a reading, no notion of an explanation, and no measure of trust.

The book's starting observation is that the equations do not have to be rewritten to answer these questions. They have to be read differently. A balance says that certain flows sum to zero at a node; a closure says that a flow and two potentials satisfy a relation; a boundary condition says that a quantity has a value. Each is a statement about which combinations of the variables are admissible. Attach to each a statement of how far from admissible the system may be, attach to each uncertain input a statement of what is known about it, attach to each sensor a statement of what it read and how well, and the collection is a probability distribution over the entire state of the system, built from nothing but the physics and the instruments. Every question in the list is then a question about that distribution, and Parts II to IV of this book are about answering them.

0.2 Where the graph comes from

A textbook on probabilistic graphical models begins with random variables and their joint distributions, defines directed and undirected graphs and the factorizations they encode, develops conditional independence, and then teaches inference on the graph. The examples are alarms and burglaries, diseases and symptoms, letters and their noisy transmissions. It is a coherent subject and the machinery is powerful, but an engineer arriving from the physical side asks a question the textbook does not answer: where did the graph come from? For an alarm and a burglary the graph is a modeling decision, drawn from a story about what causes what. For a cooling coil, a rack row or a chilled-water plant it is not a decision at all.

For a physical model the graph is the sparsity pattern of the equations. Each balance reads the flows at one node. Each closure reads one flow and the potentials at its ends. Each boundary condition reads one variable. Draw a node for every variable and a node for every equation,

balance at the heat sink:
$$Q_{2a} - Q_{12} = 0$$
the thermal interface:
$$Q_{12} - k\,(T_1 - T_2) = 0$$
the fans, to the inlet air:
$$Q_{2a} - hA\,(T_2 - T_a) = 0$$
a sensor on the heat sink:
$$T_2 = 72.0 \pm 0.5$$

read as a graph →

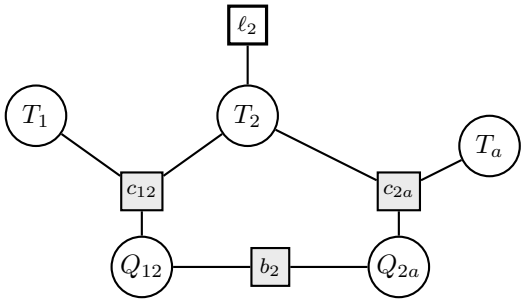


Figure 0.1: The graph comes from the equations. Left: three physics equations of the processor's heat path of Example 1.1 and one sensor reading, as an engineer writes them. Right: the same four statements as a factor graph. Each equation is a square joined to exactly the variables it reads; the sensor is a square of degree one. Nothing was decided in drawing it.

Table 0.1: What each object of a physical model becomes when probability is attached. The graph is unchanged; only the meaning of the squares is.

In the equations	In the factor graph	With probability attached
balance at a node	a factor on the node's flows	conservation, to within a small width
constitutive relation	a factor on a flow and two potentials	closure, to within its model error
boundary condition	a factor of degree one	a prior on the boundary value
uncertain parameter	a variable in every closure reading it	a prior on the parameter
sensor	(absent)	a likelihood of degree one on its variable
component in or out of service	(absent)	a discrete variable in its closure's scope

join each equation to the variables it reads, and the result is a factor graph, the bipartite graph on which the probabilistic machinery operates (Figure 0.1). No modeling decision was made in drawing it, because every edge was already an entry of the Jacobian that the solver assembles. The graph is the model's structure, seen.

Attaching probability then changes what the squares mean and not where they are. A balance equation becomes a factor that says the balance holds to within a small tolerance. A closure becomes a factor that says the constitutive law holds to within its model error. An uncertain parameter or boundary condition gets a prior, a square of degree one on its variable. A sensor becomes a likelihood, another square of degree one, on the variable it reads. A component that may be in service or out becomes a discrete variable in the scope of the closure it governs. Table 0.1 lists the correspondence, which Chapter 3 develops. The bridge from the engineer's equations to the graphical model is this table, and it is short.

0.3 One subject, not four

Little in this construction is new in isolation, and it would be wrong to present it as if it were. Four traditions have each built part of it, and the book is written in the conviction that they are one subject that has been taught as four.

Port-based physical modeling. Bond graphs, network thermodynamics and port-Hamiltonian systems supply the domain-independent view of a physical system as

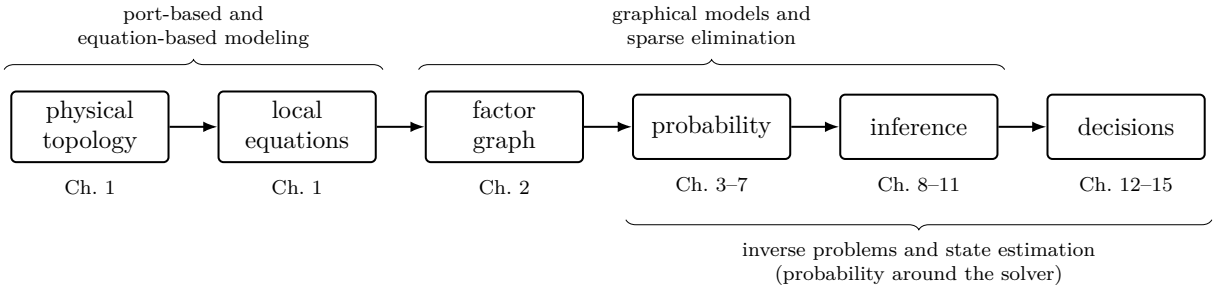


Figure 0.2: The chain of ideas the book follows, with the chapters that cover each link and the traditions that own parts of it. No one tradition carries the object from the topology to the decision.

conserved quantities at nodes, flows driven by potentials along edges, and conjugate flow-and-potential pairs at ports, the same structure whether the domain is thermal, fluid, electrical or mechanical [43, 76, 79, 104]. This tradition stops at the equations: it says how to write $\mathbf{F}(\mathbf{z}) = \mathbf{0}$ for any physical system, and then solves.

Equation-based modeling. Acausal modeling languages and the structural analysis of differential-algebraic systems establish that a physical law is a relation among variables, not an assignment of an output from inputs, and they give the bipartite equation-variable structure and its matching theory [6, 22, 72, 78]. This tradition, too, stops at the solve, and it does not attach probability.

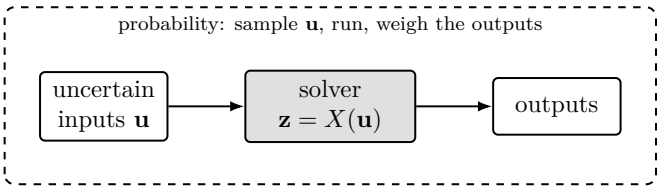
Probabilistic graphical models and sparse elimination. Factor graphs, Markov properties, separators, elimination, fill and treewidth come from the graphical-model and sparse-matrix literatures [17, 50, 54, 56, 87]. This tradition starts from a factorized distribution and asks how to compute with it; it does not say where the factorization comes from, and its graphs are drawn, not derived.

Bayesian inverse problems and state estimation. Priors on uncertain physical inputs, sensors as likelihoods, weighted least-squares estimation under conservation constraints, and the diagnosis of bad data are established in inverse problems, in power-system state estimation, and in process data reconciliation [1, 41, 65, 73, 98], and in data centers they meet the thermal mapping of a hall from its own sensors [74]. This tradition does attach probability to physics, but usually around the solver: the physics is a black-box map from inputs to observables, and the posterior is over the inputs alone.

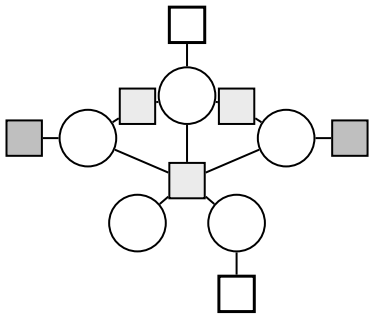
Put the four end to end and they make one chain,

$$\boxed{\text{physical topology} \;\rightarrow\; \text{local equations} \;\rightarrow\; \text{factorization} \;\rightarrow\; \text{probability} \;\rightarrow\; \text{inference} \;\rightarrow\; \text{engineering decisions}} \tag{0.2}$$

of which each tradition owns one or two links and none carries the whole. Figure 0.2 draws it, with the chapters of this book placed along it. The book's purpose is to make the chain feel like one subject: to start where the engineer starts, at the topology and the equations, and to arrive where the operator needs to be, at a decision with its uncertainty stated, without changing the object along the way.



(a) probability around the solver



(b) probability in the structure:
priors (dark) and sensors (open) added to the physics graph

Figure 0.3: Two ways to combine physics and probability. (a) The solver is a black box; probability lives outside it, over its inputs and outputs. (b) The physics equations are the factors; priors and sensors are squares added to the same graph. The book takes the second route and says why.

0.4 Probability in the structure, not around the solver

There is an obvious way to combine a physical model with probability, and most of the fourth tradition uses it: leave the solver as it is, treat it as a function from uncertain inputs to observable outputs, draw inputs from their priors, run the solver, and reason about the outputs. The physics is a black box inside a probabilistic wrapper. Figure 0.3 draws it on the left. It is correct, it is general, and it is how a great deal of uncertainty quantification is done.

The book takes the other route, drawn on the right: the equations themselves are the factors, and probability lives inside the structure, on every variable the physics has. The thesis of the book, in one sentence, is

> Do not put probability around the solver. Put it into the structure of the physical model. (0.3)

Three consequences follow, and they are what the later chapters are about.

There is one object. The deterministic solve is the special case of the probabilistic model in which the priors and sensors are absent and the physics is exact; the two share every row and every Jacobian (Chapter 6 proves this). Adding a sensor adds a row. Adding a question adds nothing: the estimate, the diagnosis, the counterfactual and the risk are readouts of one posterior (Chapter 12).

Structure is preserved, and structure is what computation costs. The wrapper's map from inputs to outputs is dense: every output depends on every input. The physics graph is sparse: every equation reads a few variables. The posterior built in the structure has the sparsity of the physics, and every computation in Part III is affordable because of it. The wrapper's route discards the sparsity the moment it closes the box.

Violation can be seen. A wrapper enforces the physics exactly, because the solver does. When the sensors report something no state of the physics can produce, the wrapper has no answer. The structural model gives each law a width, lets it be violated at a cost, and reports the violation as a diagnosis: the recirculating exhaust, the misplaced sensor, the missing blanking panel (Chapters 4 and 12).

0.5 The boundary

Every system worth modeling is a piece cut out of a larger one. A server trades air and power with its rack, a rack with its row, a contained aisle with the hall, and the hall with the chilled-water plant and the utility feed. The piece has an inside that the modeler describes and a boundary across which conserved quantities pass to an outside the modeler does not describe. Chapter 1 calls such a piece a *boundary-exchanging system*, and the book returns to the boundary at every level for a reason that is easiest to state in the language of the third tradition.

In a graphical-model text, a *separator* is a set of variables whose knowledge makes two parts of the graph independent, and finding one is a graph-theoretic exercise whose reward is a cheaper computation. Data center designers drew their separators before any theory asked them to. A rack exchanges air with the room only through its front and back faces, and power only through its distribution units; a contained aisle exchanges air with the hall only through the racks and the cooling units that bound it. The airflows and temperatures at those faces, a flow and its conjugate potential at each port, are exactly the variables that separate a rack's or an aisle's interior from the rest of the hall (Chapter 5 proves it, and proves that nothing smaller does). So the instruction "find a separator" becomes "cut the hall at the rack faces and the containment", which is where the designer has already cut it, and the mathematics then explains why that cut is also the cheap place to divide the computation. A missing blanking panel is, in this reading, a hole in the separator.

This gives one decomposition three readings at once. Physically, a subsystem is summarized to its neighbors by what crosses its ports, exactly, because conservation says the cut flows equal the net source inside. Probabilistically, the ports are a separator, so a subsystem's message to the rest of the system is a function of the ports alone. Computationally, eliminating the interior onto the ports is the Schur complement of sparse elimination, the Kron reduction of network theory, and the region message of a junction tree, which are one operation (Chapter 8). A region can be an equivalent network with error bars, computed once and reused; a large system can be a tree of regions (Chapter 17); and a partition can be judged by its ports (Chapter 18). The three-way connection is the part of the book for which the author has found the least direct precedent, and it is the reason the boundary is in the title.

0.6 From "what is the state" to "what can I know"

The pipeline (0.1) has a direction: inputs in, outputs out. A physical law does not. The relation

$$Q - UA\,(T_h - T_c) \;=\; 0 \tag{0.4}$$

says that four quantities satisfy one constraint. It does not say that Q is computed from the temperatures; given Q and T_h it determines T_c, and given all three it determines UA, which is how a fouled cooling coil is found (§12.8). Which variable is unknown is a property of the question, not of the physics (Chapter 2 makes this precise, and shows that for a looped system no fixed direction exists at all). The book keeps the relations as relations, which is what the second tradition insists on and what the first tradition's bond graphs draw, and it is why the factor graph rather than the directed graph is the book's language.

Once uncertainty and observations enter, the absence of direction becomes the point. A reading of one temperature is evidence about everything the physics ties to it. On the processor's heat path that runs through the book, a single thermometer on the heat sink cuts the uncertainty

in the processor's unmetered power by a factor of three and a half, and predicts the unmeasured package temperature to within a degree and a half, before any second sensor is read (Chapter 3). That prediction is a *virtual sensor*, and its error bar is part of the answer. The question the model answers has changed from "what is the state, given the inputs" to "what can I know, given what I have", and the second question is the operator's.

0.7 What is inherited and what is the book's

A book that combines old ingredients owes its reader a clear account of which is which. Each chapter closes with a section of notes that does this for its own contents. The account for the whole is short.

Inherited: the port-based view of physics; the acausal reading of its equations; the factor graph, its Markov properties, elimination and treewidth; the Gaussian machinery of least squares, sparse Cholesky and the Laplace approximation; Bayesian inverse problems; state estimation and data reconciliation with their bad-data statistics; the causal calculus of interventions; variance-based attribution; reliability evaluation and failure-set analysis. None of these is the book's, and the notes say whose each is.

The book's: the synthesis. Taking an equation-based, conservation-structured physical model and turning its residual structure, unchanged, into the factorization on which inference, diagnosis, uncertainty propagation, attribution and decision queries are performed; the identification of physical ports with probabilistic separators and computational cuts; the intervention semantics for acausal equations; and the insistence, carried through every chapter, that the deterministic model is a corner of the probabilistic one.

The author has looked for this treatment and not found it. Pieces of it exist in several communities that have discovered the connection independently. Multibody dynamics has been written as a factor graph whose sparse structure serves both simulation and estimation, with forward and inverse dynamics as different queries on one graph [11], which is the "relations, not assignments" reading in a single domain. Bond graphs have been converted into Bayesian networks for fault diagnosis [62], which attaches probability to a port-based physical model but does so by assigning the causal directions this book declines to assign. Power systems have been estimated by message passing on a factor graph built from the power-flow equations [15], which is Chapters 6 to 12 in one domain and one query. Table 0.2 sets these beside the four traditions and beside this book. What the table shows is not that any column is empty but that no prior row fills them all, and that several communities have arrived at parts of one construction without a text that says they are parts of one construction. That is the gap the book is written to fill, and "not found" is the strongest claim the author is willing to make for it.

0.8 How to read this book

Who it is for. The reader is an engineer who knows conservation laws and linear algebra and has never met a probabilistic graphical model. Nothing about probability beyond the meaning of a density is assumed; Chapter 3 supplies what is needed in a page and Appendix B the Gaussian identities. A reader who knows graphical models and wants the physics will find Chapters 1, 2 and 5 the ones to read first, since they say what the graph is and why the ports separate. The examples are data center examples; the constructions are not, and a reader who works on water,

Table 0.2: The closest prior work, by what it covers. A check marks a substantial treatment; a word marks a partial or specialized one.

	physics graph	equations as factors	probability	general domains	ports as separators	broad inference
bond graphs, port-Hamiltonian [79, 104]	✓	—	—	✓	✓	—
multibody factor graphs [11]	✓	✓	limited	multibody	—	solve, estimate
bond-graph Bayesian networks [62]	✓	indirect	✓	general idea	—	fault diagnosis
factor-graph state estimation [15]	power grid	specialized	✓	power only	regional	state estimation
graphical-model texts [50, 56]	—	—	✓	abstract	graph-theoretic	✓
inverse-problem texts [41, 98]	equations	not generally	✓	✓	—	inverse only
this book	✓	✓	✓	✓	✓	✓

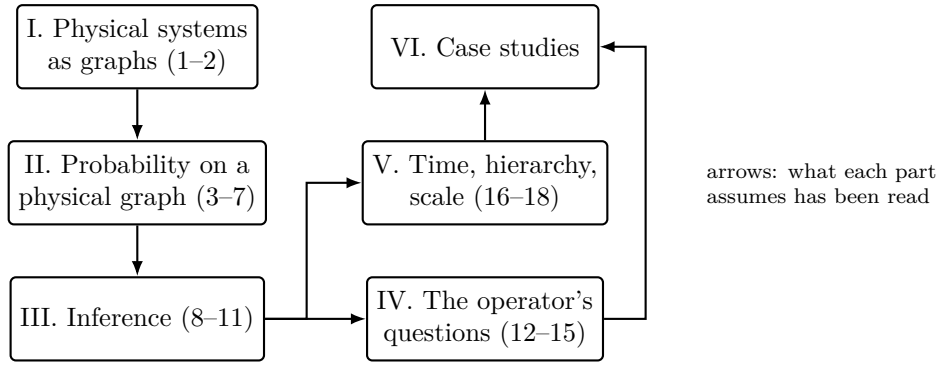


Figure 0.4: The parts of the book and their dependencies. Parts IV and V each need Part III and not each other; Part VI needs both.

district heating or any other conserved quantity will find nothing in Parts I to V that needs changing but the names.

The parts. Figure 0.4 shows the six parts and what each depends on. Part I is deterministic and builds the objects. Part II attaches probability and states, rather than skips, the places where that attachment needs care. Part III organizes the computation. Part IV asks the operator's questions. Part V handles time, hierarchy and scale. Part VI is a set of complete studies, each reproducible from the material in the earlier parts.

The running examples. Two small systems appear in almost every chapter: the processor's heat path of Example 1.1, from the package through the thermal interface to a heat sink that the server's fans cool with cold-aisle air, and the facility's own distribution, a UPS busbar feeding two units in a triangle under the DC approximation of power flow. Both are introduced in Chapter 1, both are small enough to solve by hand, and both acquire priors, sensors, a failed component, a second region, an intervention and a failure set as the chapters go on. The reader who follows them through will have seen every construction of the book on a system whose every number can be checked. Larger systems appear where the small ones cannot show the point: the facility's six-bus distribution for decomposition and failure, a cooled room for common cause, a conducting plate for cost at scale, and the case studies of Part VI.

The numbers. Every number in every worked example is produced by a short script that accompanies the book, one per chapter, and the text says which. The scripts are a few hundred lines each and use nothing beyond a numerical library; they are meant to be read and rerun, and several exercises ask for exactly that. The case studies of Part VI read their numbers from committed result files of the studies they describe. No number in the book was typed by hand.

Conventions. Each chapter ends with a section on what is in practice, notes on what is inherited and from where, a summary, five exercises, and worked solutions to two of them. Load-bearing conclusions are set apart as numbered principles; there are about three per chapter, and a reader who remembers only those will have the book's argument. Propositions are proved where they are stated. Notation is collected in Appendix A, and the reference implementation that the author used to run every script is mapped to the book's constructions in Appendix C; nothing in the body depends on it.

0.9 What this book does not do

It does not teach physical modeling; it assumes the reader can write a balance and a closure, and it uses the simplest ones. It does not teach probability or graphical models in generality; it teaches the parts a physical model needs, and points to the texts that teach the rest. It does not develop machine learning: no factor in this book is learned from data, though Chapter 12 estimates parameters and Chapter 14 fits a surrogate, and the notes say where the learned versions of its constructions live. It does not treat airflow as a field: there is no computational fluid dynamics here, and a hall's air is a network of lumped nodes and conductances whose parameters a field model, where one exists, would supply. It does not schedule the IT workload; loads are inputs with priors, not decisions the book optimizes. It does not treat continuous space: every system here is a graph of lumped elements, and a field discretized on a mesh appears only as a large graph whose treewidth is a warning. And it does not claim that the constructions it teaches are new; it claims that they are one subject, and it tries to teach them as one.

Notes and further reading

The five gaps that §0.1 to §0.6 describe were articulated in a review of the manuscript, and the positioning of §0.7 follows a literature search made in response to it. The closest works found, Blanco-Claraco, Leanza and Reina's factor-graph multibody dynamics [11], Lo, Wong and Rad's bond-graph Bayesian networks [62], and Chavali and Nehorai's factor-graph state estimation [15], are each discussed in the chapter whose subject they touch. The four traditions of §0.3 are cited there and again in the notes of the chapters that inherit from them; the reader who wants one text per tradition is directed to Karnopp and Rosenberg [43], Benveniste, Caillaud and Malandain [6], Koller and Friedman [50], and Kaipio and Somersalo [41].

Part I

Physical systems as graphs

Chapter 1

Conservation, closure, and the boundary

Every physical system a modeler cares about is a piece cut out of a larger one. A building trades heat with the weather. An electrical room trades power with the rest of the facility. A rack trades air with the room. The piece has an inside, which the modeler describes in detail, and a boundary, across which conserved quantities pass to an outside the modeler does not describe at all.

This chapter builds the deterministic model of such a piece. It has three ingredients and no probability. The three ingredients are conservation at places, closure along paths, and exchange at the boundary. The reason to build the deterministic model first, and to build it carefully, is that the probabilistic model of Part II inherits its structure entirely: the places, paths and boundary of this chapter become the nodes, factors and separators of the graphical model. Nothing is added to the structure later. Only uncertainty is.

1.1 Three questions about any system

Take any conserved quantity: energy, mass, electric charge, a chemical species. Three questions describe how it lives in a system.

1. Where can it accumulate?
2. How does it move from one place to another?
3. What crosses the boundary?

The first question identifies the *places*. Each place has an amount of the quantity, which may change in time. The amount is an *extensive* state: the energy in a wall, the mass of water in a tank, the charge on a capacitor. Doubling the place doubles the state.

The second question identifies the *paths*. Along a path the quantity flows at some rate: watts, kilograms per second, amperes. What drives the flow is a difference in an *intensive* quantity between the two ends of the path: temperature, pressure, voltage. Doubling the place does not double the intensive quantity; it is a property of the state at a point, not of the amount.

The third question identifies the *boundary*. Some paths lead to places the model does not describe. The quantity that crosses those paths enters or leaves the system. The places on the far side are held at imposed conditions, either an imposed intensive value (the ambient temperature, the grid voltage) or an imposed flow (a heat load, a scheduled injection).

Table 1.1: Four conservation domains and their three quantities.

Domain	Extensive state	Flow	Intensive driver
Energy (thermal)	internal energy E [J]	heat rate Q [W]	temperature T [K]
Mass	mass M [kg]	mass rate $\dot{m}$ [kg/s]	pressure p [Pa]
Charge	charge q [C]	current I [A]	voltage V [V]
Power (DC network)	—	line flow F [W]	angle θ [rad]

Table 1.1 lists the four conservation domains that recur in this book with their extensive state, their flow, and the intensive quantity that drives it. The rows are physically different; the columns are the same three questions. That parallel is the foundation of everything that follows. A method that works on the columns works on every row.

The last row is worth a comment. In the DC approximation of a power network the conserved quantity is real power, the flow on a line is the power it carries, and the driver is the voltage angle difference between its ends. There is no extensive state because the approximation is steady: buses do not store power. The row still fits the three questions. The first answer is simply "nowhere".

1.2 Nodes conserve

The three questions define a graph. The places are nodes; the paths are edges. Write $\mathcal{N}$ for the set of nodes, $\mathcal{E}$ for the set of edges, and $\mathcal{D}$ for the set of conservation domains the model tracks. A model with heat and water flowing through the same pipework has $\mathcal{D} = \{\text{energy}, \text{mass}\}$; a DC power network has $\mathcal{D} = \{\text{power}\}$.

1.2.1 The balance law

At each node n and for each domain d in which n can accumulate, the rate of change of the extensive state equals what flows in, minus what flows out, plus what is generated, minus what is lost:

$$\boxed{\frac{\mathrm{d}X_{n,d}}{\mathrm{d}t} \;=\; \sum_{e \in \mathcal{E}(n)} A_{n,e}\, P_{d,e} \;+\; G_{n,d} \;-\; L_{n,d}} \tag{1.1}$$

Here $X_{n,d}$ is the extensive state of domain d at node n, $\mathcal{E}(n)$ is the set of edges incident on n, $P_{d,e}$ is the flow of d along edge e, $G_{n,d}$ is a source at the node and $L_{n,d}$ a loss. The coefficient $A_{n,e}$ is the sign that says whether edge e delivers to or draws from node n; it is defined in §1.2.3.

Equation (1.1) is the whole of physics that a node contributes. It is linear in the flows, always and exactly. It has no parameters. Whether the domain is energy or mass or charge, whether the node is a wall or a tank or a bus, the balance row looks the same. This is not a modeling convenience. It is the reason the graph is the right object: the graph is the conservation skeleton, and every domain that the model tracks uses that same skeleton.

1.2.2 Node roles

Not every node accumulates. Three roles cover the cases.

Balanced. The node has an extensive state that can change. Equation (1.1) holds with its left side live. A wall, a water tank, a battery.

Zero-storage. The node is a junction or a bus. It has no extensive state, so equation (1.1) holds with $\mathrm{d}X_{n,d}/\mathrm{d}t \equiv 0$: what comes in goes out. A pipe tee, an electrical bus, a mixing point.

Reservoir. The node is on the far side of the boundary. Its intensive state is imposed, not solved. It contributes no balance row, because the model does not claim to know its balance; it appears only as the far end of the edges that cross the boundary. The atmosphere, the utility grid, a feed line.

The first two roles are the inside of the system. The third is the outside. Which nodes are reservoirs is the modeler's first and most consequential decision, and §1.4 returns to it.

A node also declares which domains it is balanced in. A pipe carrying water may need a mass balance at each tee but no energy balance if the pipe is adiabatic; a bus needs a power balance and nothing else. The balance rows exist only where the modeler asks for them. The roles are declared per domain as well: a return manifold whose pressure is held by a pump is a reservoir for mass and yet a balanced node for energy, because the temperature of the water leaving it is the model's to determine. Example 1.3 has such a node.

1.2.3 Incidence

Give every edge a direction, from a tail node to a head node. The direction is a bookkeeping convention, not a physical claim: a flow that runs against the edge's direction is simply negative. The *incidence matrix* $\mathbf{A}$ has one row per non-reservoir node and one column per edge, with entries

$$A_{n,e} = \begin{cases} +1 & \text{edge } e \text{ enters } n \text{ (n is its head)}, \\ -1 & \text{edge } e \text{ leaves } n \text{ (n is its tail)}, \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

Every column of the full incidence matrix (with reservoir rows included) has one $+1$ and one -1, so the columns sum to zero. Dropping the reservoir rows breaks that property for the columns of boundary edges, and that is exactly right: the flow on a boundary edge is not balanced by any row of the model. It is what the system exchanges.

With $\mathbf{A}$ in hand, the sum in equation (1.1) is one row of a matrix product. Stack the flows of domain d into a vector $\mathbf{P}_d$ and the balances into a vector; the balance law for all nodes at once is $\mathrm{d}\mathbf{X}_d/\mathrm{d}t = \mathbf{A}\,\mathbf{P}_d + \mathbf{G}_d - \mathbf{L}_d$.

> **Principle 1.1.** *The incidence matrix is shared by every domain. Energy and mass in the same pipework do not have two graphs. They have one graph, and each domain contributes balance rows on the nodes where it is tracked, using the same* $\mathbf{A}$.

An edge that carries one domain but not another, such as a rotating shaft that transmits energy but no mass, or a conduction path through a solid, simply has no flow variable for the absent domain. Its column contributes zero to that domain's balance rows.

1.3 Edges close

The balance law says that the flows at a node sum to the accumulation. It does not say what any flow is. That is the job of the edge.

Table 1.2: Four closure patterns. Each is written as a residual that vanishes when the relation holds.

Pattern	Residual	Typical use
Carrier flow	$P_e - q_e\,\varepsilon_e$	a quantity carried by another: $P = \dot{m}\,h$ (enthalpy carried by mass), $P = I\,V$ (power carried by current)
Gradient flow	$P_e - g(\phi_i, \phi_j; \theta)$	driven by a difference in intensive state: $Q = UA\,(T_h - T_c)$, $F = B\,(\theta_i - \theta_j)$
Property	$y - f(x_1, \ldots, x_k)$	a property table or a device specification: $h = h(p, T)$, $Q = \mathrm{COP} \cdot W$
Piecewise	$y - f_b(x)$	a device that changes regime: a valve that saturates, a limiter, a breaker

1.3.1 Closure relations

Along each edge, for each domain it carries, there is a flow variable and a relation that ties that flow to the states at the edge's ends and to some parameters. The relation is called a *closure* because it closes the system of equations: the balance rows alone have more unknowns than rows, and the closures supply the missing rows.

Definition 1.1 (Closure)**.** A closure is a local relation $r(\mathbf{z}_{\mathrm{loc}}; \theta) = 0$ among a small set of variables $\mathbf{z}_{\mathrm{loc}}$, with parameters θ, written in *residual form*: a function that is zero exactly when the relation holds. The closure declares which variables it reads.

Four patterns cover the great majority of physical closures (Table 1.2).

The gradient pattern is the one the three questions anticipated: a flow driven by an intensive difference. Conduction, convection, DC current, DC power flow and laminar pipe flow all take this form, with different functions g and different parameters θ: a conductance, a heat transfer coefficient times an area, a susceptance. The carrier pattern appears whenever one domain rides on another, as enthalpy rides on mass flow. The property pattern is a catch-all for relations among variables at a single place. The piecewise pattern is the first hint of something the deterministic model handles awkwardly and the probabilistic model handles naturally; it returns in Chapter 7. Appendix D catalogues the closures of the domains this book touches. For each it gives the residual and its pattern, the unknowns and parameters, the Jacobian, where the relation kinks, its shape, and the parameter that usually carries the prior.

1.3.2 A closure is a relation, not an assignment

Write the conduction closure as most textbooks do:

$$Q = UA\,(T_h - T_c). \tag{1.3}$$

The notation suggests a computation: given the two temperatures, compute the heat rate. But nothing in the physics says which quantities are given. If the heat rate and the hot temperature are known, the same relation determines the cold temperature:

$$T_c = T_h - Q/(UA). \tag{1.4}$$

If all three are measured, the relation determines the conductance UA. The physics asserts that four quantities satisfy one constraint. It does not assert a direction.

The residual form makes this explicit:

$$r(Q, T_h, T_c; UA) \;=\; Q - UA\,(T_h - T_c) \;=\; 0. \tag{1.5}$$

Equation (1.5) has no left-hand side to compute. It is a surface in the space of its variables, and every point on the surface is a physically admissible state of the edge.

> **Principle 1.2.** *A closure relates its variables. It does not compute one of them from the others. Which variable is unknown is a property of the question being asked, not of the physics.*

This principle looks like a remark about notation. It is not. It decides, in Chapter 2, which kind of graphical model a physical system should be written in, and it is the reason the arrows of a Bayesian network are not the natural representation of a closure. Hold on to it. The principle itself is old: it is the acausal view of physical laws that bond-graph modeling and equation-based modeling languages are built on [72, 79], and §1.8 says more.

1.3.3 Parameters

The parameters θ of a closure, a conductance, a susceptance, an efficiency, are ordinarily numbers the modeler supplies. Nothing forces that. A parameter the modeler does not know can be left as an unknown, added to the vector of variables, and determined by the equations if enough other things are known. A conductance inferred from a measured heat rate and two measured temperatures is the simplest case of this. The deterministic model needs one extra equation for each parameter it frees; the probabilistic model of Part II needs only a prior.

1.4 The boundary

The boundary is the set of edges with a reservoir at one end. Everything on the reservoir's side is outside the model. This section says what the model knows about the outside, why that knowledge is exact in one specific sense, and why the boundary is the most important part of the system.

1.4.1 Reservoirs and imposed conditions

A reservoir contributes no balance row. It contributes an imposed condition on the boundary edge, of one of two kinds.

Imposed potential. The reservoir's intensive state is fixed: the ambient temperature, the grid voltage angle at the point of connection. The boundary edge's closure then determines the flow across it. This is the Dirichlet condition of boundary-value problems.

Imposed flow. The flow across the boundary edge is fixed: a specified heat load, a scheduled generation injection. The balance row at the inside node then determines the inside potential. This is the Neumann condition.

Each boundary edge is thus a pair, a flow and its conjugate potential, one of which is imposed and the other of which the model determines. The pair is domain-independent (Table 1.3). Call the pair a *port*.

END OF THE PREVIEW

Probabilistic Modeling of Data Centers

These are the opening pages. The complete book runs 390 pages.

The full PDF opens with a token. Write to **rmarquez123@gmail.com**, name the book, and I will send you one.

One token opens every book on the site.

Ricardo Marquez, Ph.D.

ricardo-marquez.com