Conductance dimensioning and the reliability of conclusions drawn from graph-based thermal inference in data centers

Ricardo Marquez, Ph.D.
rmarquez123@gmail.com

September 2026

Abstract

A data center's thermal behavior is often modeled as a graph of lumped links: racks, rows, rooms and a water loop, each pair joined by a thermal conductance. Those conductances are a modeling choice, and this report asks how far the conclusions of a collection of such models depend on it. It is a sensitivity analysis over that choice, and the experiment that explains its findings is an inference study, in which unknown heat terms are estimated from temperature readings. Forty-five models of one collection, spanning five levels from a whole building to a single chassis, are compared under two dimensionings: a low-conductance condition, with rack links of 10 to 50 W/K, and a sized condition, in which every link carries its design heat at a 12 K rack air rise and 3 K coil approaches. Under the low-conductance dimensioning 18 of 45 models keep every solved node at or below the 320 K air limit, and 18 do not, the hottest solving at 6,134 K; under the sized dimensioning 32 do, and the remainder are the 9 models that solve no thermal network and 4 whose hot states are reported on purpose. Re-deriving the 18 affected conclusions leaves 3 unchanged and 7 with the same direction and a different magnitude; 1 changes its framing, 2 reverse, and 5 were wrong independently of the dimensioning. A controlled experiment on one three-room building shows why diagnostic evidence shrinks: the evidence for a fault follows the part of its sensitivity that competing faults cannot explain, which gives ratios of 0.2167, 0.07773 and 0.03434 at the rack, room and building levels, reproduced by a closed form to within 10^{-3} relative. The whole comparison costs 45 model runs and one pair of building solves per level.

1 Introduction

Data centers use a small but significant share of the world's electricity, and a large part of that share is spent moving heat from silicon to the outside air. Masanet, Shehabi, Lei, Smith and Koomey (2020) put data-center electricity use at about one percent of global demand and attribute its slow growth, while computing grew many times over, to efficiency gains that include cooling; Barroso, Hölzle and Ranganathan (2019) describe how warehouse-scale facilities are cooled, and Ebrahimi, Jones and Fleischer (2014) review the cooling technologies and the operating temperatures they allow. Whether a facility is safe, how much capacity it can use and where a fault sits are all questions about temperature, and answering them needs a model of how heat travels from a rack to the chilled water. Computational fluid dynamics resolves the air, at a cost that rules it out for repeated what-if questions; the alternative is a compact model, a graph whose nodes carry temperatures and whose links carry heat in proportion to a temperature difference. Fulpagare and Bhargav (2015) review these compact models and their use in control and diagnosis. The proportionality constant of each link, its thermal conductance, is where the physics of the air stream enters such a model, and it is set by the designer rather than measured.

Estimating what a network is doing from a few readings on it is an old construction. Schweppe and Wildes (1970) framed the state of a power system as a weighted least-squares fit to redundant measurements, the method still used to monitor transmission grids; factor graphs (Kschischang, Frey and Loeliger 2001) write such problems as sparse products of local relations; and where a fit is used to decide between explanations, the quality of the decision depends on which quantities the readings can separate, a question of identifiability that Raue et al. (2009) treat through the profile likelihood. Where to put sensors so that they separate as much as possible is itself a studied problem, with greedy selection under a variance-reduction objective as the standard approach (Krause, Singh and Guestrin 2008). None of these constructions is new here. What is asked here instead is a question of verification in the sense of Oberkampf and Roy (2010): whether the conclusions such a model produces survive a change in one of its inputs that the modeler chose. Saltelli et al. (2007) set out how such a question is posed: a sensitivity analysis states which inputs are varied, over which range, and what output is watched, and it is only as broad as the set of choices it varies, which is a reason to vary a modeling choice across a whole collection of models rather than one.

TERRA is a general-purpose engine for probabilistic inference on physical networks. It treats a network as a graph on which the physics is written locally: a conservation law at every node, a constitutive law on every edge, and, attached to the same graph, whatever the sensors report and whatever is known in advance. Each of these is a relation that holds only to within a stated tolerance, and the engine weighs them together to find the most probable state of the system, the uncertainty of every quantity in it, and the relative plausibility of competing explanations of the same readings. Domains enter through plugins that supply their laws, units and file formats; the inference and the bookkeeping are shared. This collection of models is written with its data-center plugin, which supplies the heat balances of racks, rows, rooms and loops, the sizing helper that turns a design heat into a conductance, and the fault terms that a diagnosis estimates; the engine's fit, its uncertainties and its comparisons of competing explanations are what the collection's conclusions rest on. Section 3 states the framework and each of these constructs precisely.

This report takes forty-five such models, from a single chassis to a whole building, and asks three questions about the conductances they carry. First, whether their solved temperatures are physical: how many of them keep every solved node at or below the 320 K air limit under each of two dimensionings, and what the remainder are. Second, whether their conclusions survive: which of the conclusions that the dimensioning can move keep their direction, which change magnitude only, and which reverse. Third, why diagnostic evidence changes as it does, which a controlled experiment on one building answers with a closed form. Section 2 describes the collection and the two dimensionings, Section 3 the model, the sizing rule, the re-derivation protocol and the metrics, Section 4 the results, Section 5 their limits and Section 6 concludes.

2 System and data

The physical system is the one drawn in Figure 1. Electrical power dissipated in a rack leaves it as heat in the air stream crossing it; the row's legs carry that air to and from the room; the room's coil passes the heat to a building water loop, and the loop passes it to chilled water held at a fixed temperature. Each of those stages is one link of a graph, and each link carries heat in proportion to the temperature difference across it. A rack also leaks: its electronics draw more power as they warm, which the models represent as a slope α in watts per kelvin above a reference temperature.

The collection holds forty-five models of such systems, grouped by the level they describe: 19 at the building level, 14 at the room or hall level, 6 at the row or pod level, 5 at the rack level, 1 at

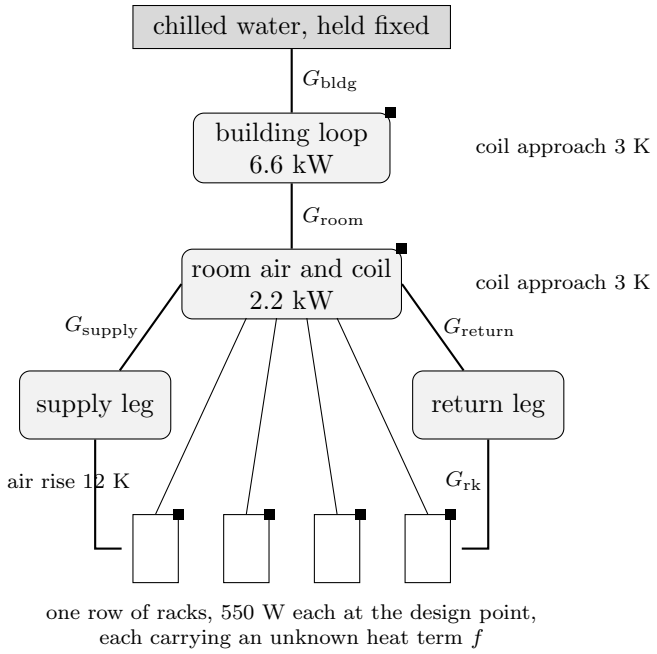


Figure 1: The physical system every model in the collection is a part of. Electrical power dissipated in a rack leaves it in the air stream that crosses it, warming that air by a design rise of 12 K; the row's supply and return legs carry the air between the racks and the room; the room's cooling coil passes the heat to the building's water loop across a design approach of 3 K, and the building loop passes it to chilled water held at a fixed temperature. Each stage is one link with a thermal conductance G. Black squares are temperature sensors: a model may have them at racks, at room air and on the loop. Design heats are those of the three-room building used for the mechanism experiment of Section 3.6: 550 W per rack, 2.2 kW per room and 6.6 kW for the building. Every rack, room and building carries an unknown heat term f, the quantity a diagnosis estimates.

the internal component level. They answer different questions — capacity, containment, control, fault diagnosis, sensor placement, forecasting, the cost of a solve — and thirty-six of them solve a thermal network; the remaining nine work on linearized state-space models or on plant laws and solve no network, so they carry no temperature this report can screen. Table 4 in the appendix lists every model. No measured data enter any of them: every load, boundary temperature and sensor reading is generated by the model itself, and the readings that the diagnosis and placement models consume are true model temperatures perturbed by a deterministic, seeded noise of a stated standard deviation.

Two dimensionings of these same models are compared. In the low-conductance condition, rack links carry 10 to 50 W/K and room links 20 to 300 W/K, values that do not follow from the heat the links carry: a 550 W rack behind a 10 W/K link warms its air by 55 K. In the sized condition every link's conductance is its design heat divided by the temperature difference the design allows across it, a 12 K rise across a rack and 3 K approaches on the coils, which gives rack links of 2.5 to 54.2 W/K. Table 1 states the rule and both ranges. Figure 2 shows what a model is built from: the chain of links, and the four kinds of rows attached to it.

The assumptions common to every model of the collection are these.

- Each model is at steady state where it reports a steady state; the transient models integrate the same relations in time.

Table 1: The sizing rule, and the two dimensionings it separates. A link's conductance is the heat it carries at the design point divided by the temperature difference the design allows across it (Eqs. (19)–(21)). The last two columns give the range, over the thirteen models whose links carry a design heat, of the rack and room conductances before and after sizing; the building loop's values are model-specific and are listed per model in Table 4. A rack link of 10 W/K carrying a 550 W rack implies a 55 K air rise, which is why the low-conductance models solve far above the 320 K air limit.

link	design heat	design difference	conductance	low-conductance [W/K]	sized [W/K]
rack air stream	rack power P	rise 12 K	$P/\Delta T_{\text{rise}}$	10–50	2.5–54.2
room coil	$N_{\text{rk}}P$	approach 3 K	$N_{\text{rk}}P/\Delta T_{\text{app}}$	20–300	100–54,315
building loop	$N_{\text{room}}N_{\text{rk}}P$	approach 3 K	$N_{\text{room}}N_{\text{rk}}P/\Delta T_{\text{app}}$	—	—

- Heat flow on a link is its conductance times the temperature difference across it. The link is symmetric, which is exact for the steady balance of a fixed airflow and does not represent flow reversal.
- A rack's power is linear in its temperature through the leakage slope α.
- Sensor noise is Gaussian with a stated standard deviation, and unknown heat terms carry zero-mean Gaussian priors with a stated standard deviation per level.
- Physical relations are weighted with a small tolerance rather than enforced exactly, so a fit satisfies them to within that tolerance.
- Conductances, base powers, the leakage law and boundary temperatures are given; in a diagnosis only the heat terms are uncertain.
- Every temperature quoted from a model run is read at the screen's precision of 1 decimal, and quantities at the level of floating-point round-off are stated as bounds rather than printed with their digits.

3 Methods

3.1 The thermal network and its rows

The unknowns of a model are the temperature of every node, the heat flow on every link, the power of every rack and, where the model performs a diagnosis, an unknown heat term at every node. Table 5 in the appendix lists the symbols. Heat balances at a rack, the rack's power follows its leakage law, and the heat on a link follows the link's conductance:

$$P + f_{\text{rk}} - Q_{\text{rk}} = 0, \tag{1}$$

$$P - P_0 - \alpha\,(T_{\text{rk}} - T_{\text{ref}}) = 0, \tag{2}$$

$$Q_{\text{rk}} - G_{\text{rk}}\,(T_{\text{rk}} - T_{\text{room}}) = 0. \tag{3}$$

A room collects the heat of its racks and passes it to the building loop, and the loop passes the heat of all rooms to chilled water at the fixed temperature T_{chw}:

$$\textstyle\sum_{\text{rk}\in\text{room}} Q_{\text{rk}} + f_{\text{room}} - Q_{\text{room}} = 0, \tag{4}$$

$$Q_{\text{room}} - G_{\text{room}}\,(T_{\text{room}} - T_{\text{bldg}}) = 0, \tag{5}$$

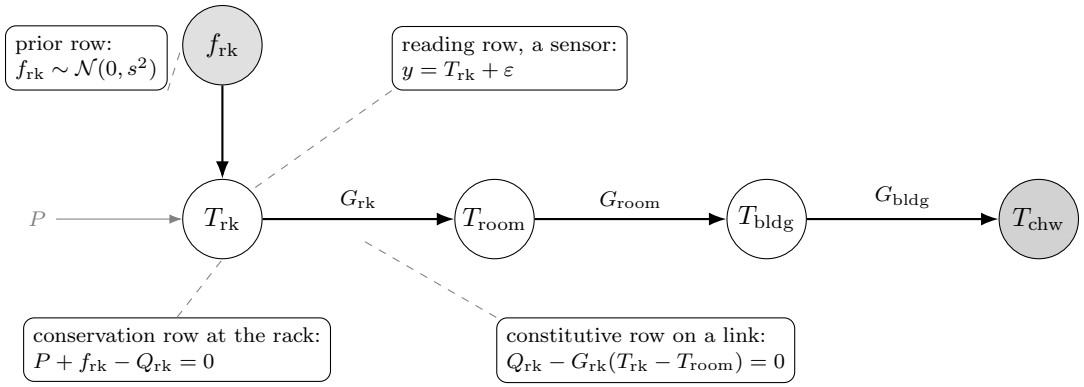


Figure 2: The model of the chain, and the four kinds of rows attached to it. White circles are unknowns that the physics determines: the rack, room and building temperatures, and the heat flows on the links. The shaded circle is an unknown heat term, the quantity a diagnosis estimates; the gray circle is the chilled-water temperature, which is given, as is the rack's electrical power P. The boxes are the rows: conservation of heat at a node, a constitutive law on each link, one reading row per sensor and one prior row per unknown heat term. Each link's conductance is the quantity this report varies; for the three-room building of Section 3.6 the sized values are 45.8, 733 and 2200 W/K against 10, 20 and 100 W/K in the low-conductance condition. One caveat: air carries heat in one direction, but each link here is a symmetric conductance, which is exact for the steady balance of a fixed airflow and does not represent flow reversal.

$$\textstyle\sum_{\mathrm{room}} Q_{\mathrm{room}} + f_{\mathrm{bldg}} - Q_{\mathrm{bldg}} = 0, \tag{6}$$

$$Q_{\mathrm{bldg}} - G_{\mathrm{bldg}}\,(T_{\mathrm{bldg}} - T_{\mathrm{chw}}) = 0. \tag{7}$$

A sensor on a temperature adds a reading row, and each unknown heat term a prior row:

$$y = T + \varepsilon, \tag{8}$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2), \tag{9}$$

$$f \sim \mathcal{N}(0, s^2). \tag{10}$$

A model with no unknown heat term and as many independent relations as unknowns has a unique solution; a model with heat terms, readings and priors is a fit, and the next subsection states what is fitted.

3.2 The probabilistic model

Every model of this collection is one instance of the following construction, whose uncertainties come from the curvature of the fit at its optimum in the manner of Tierney and Kadane (1986). A physical network is a graph, and the engine writes its physics on that graph. The state is a vector z of unknowns: one potential per node, one flow per edge, and whatever else is uncertain, such as demands or parameters. Every relation the state must satisfy is a factor, written as a residual $r_k(z)$ that vanishes when the relation holds and carrying a weight w_k, the inverse variance of how far it may be off. A conservation law at a node, a constitutive law on an edge, a sensor reading and a prior on an approximately known quantity are all factors of this one kind; they differ only in their residual and their weight, a physics relation being trusted to a small tolerance, a reading to its sensor's noise, a prior to its stated spread. Treating each factor as an independent Gaussian

belief about its residual gives the posterior density of the state given the readings y,

$$p(z \mid y) \propto \prod_k \exp\big(-\tfrac{1}{2}\, w_k\, r_k(z)^2\big). \tag{11}$$

The state that maximizes it, the *maximum-a-posteriori* (MAP) state z^*, minimizes the weighted sum of squared residuals $C(z)$ and is found by Gauss–Newton iteration on the sparse Jacobian J of the residuals, $J_{ki} = \partial r_k / \partial z_i$, with W the diagonal matrix of weights, from a flat start until the gradient norm falls below a tolerance ϵ:

$$C(z) = \tfrac{1}{2} \sum_k w_k\, r_k(z)^2, \tag{12}$$

$$z^* = \arg\min_z C(z), \tag{13}$$

$$z \leftarrow z - (J^\top W J)^{-1} J^\top W\, r(z) \quad \text{until } \|J^\top W\, r(z)\| < \epsilon. \tag{14}$$

Near the minimum the cost (12) is approximately quadratic, so the posterior is approximately Gaussian about z^* with information matrix Λ. This is the *Laplace approximation*, and the *posterior standard deviation* σ_i of any unknown is read from the diagonal of Λ^{-1}, from the same sparse factorization the last Gauss–Newton step used:

$$p(z \mid y) \approx \mathcal{N}\big(z^*, \Lambda^{-1}\big), \tag{15}$$

$$\Lambda = J^\top W J, \tag{16}$$

$$\sigma_i = \sqrt{(\Lambda^{-1})_{ii}}. \tag{17}$$

When several factor graphs compete to explain the same readings — a leak here, a leak there, no leak — each is compared by its *evidence* $Z = \int \prod_k \exp(-\frac{1}{2} w_k r_k(z)^2)\, dz$, the probability of the readings under the model once the extra freedom of its unknowns has been paid for. In the Laplace approximation the evidence has a closed form, and normalizing every factor by its own weight lets graphs with different numbers of unknowns d and factors compare as densities,

$$\log Z = -C(z^*) + \tfrac{d}{2} \log 2\pi - \tfrac{1}{2} \log\det \Lambda + \sum_k \tfrac{1}{2} \log \frac{w_k}{2\pi}. \tag{18}$$

Differences of $\log Z$ between hypotheses, from (18), are in *nats*, natural-log units of odds; five nats is odds of about 150 to 1. The posterior probability of a hypothesis is its evidence normalized over all hypotheses under equal prior odds. Everything that follows is a choice of variables and factors on this graph, followed by these three operations.

Background: the model, the fit, and its uncertainty. The quantities to be determined are the temperatures of racks, room air and water loops, the heat flows on the links between them, and the unknown heat terms a diagnosis estimates. Physics ties them together and the sensors constrain them. At every node a conserved quantity balances: the heat entering a rack, a room or the building loop equals the heat leaving it, as in Eqs. (1) and (4). Along every edge a closure law relates the flow to the potential difference across it: the heat on a link is its conductance times the temperature difference across it, as in Eq. (3). Every sensor adds one relation between its reading and the quantity it observes (a temperature sensor, Eq. (8)), and every quantity known approximately in advance adds a prior, a relation between the unknown and its expected value (an unknown heat term with a stated spread per level, Eq. (10)). Each relation is written as a residual that vanishes

when it holds, and together the residuals form one sparse system; this report calls them rows. No row is enforced exactly. Each carries a weight, the inverse variance of how far it may be off: a physics row is trusted to a small tolerance, a reading to its sensor's noise, a prior to its stated spread. The *maximum-a-posteriori* (MAP) state is the set of unknowns that minimizes the weighted sum of squared row misfits, $C(z) = \frac{1}{2}\sum_k w_k\, r_k(z)^2$; with only physics and readings it is the weighted least-squares fit, and it is found by Gauss–Newton iteration. The *posterior standard deviation* of an unknown is its uncertainty bar: near the minimum the misfit is approximately quadratic, so the state is approximately Gaussian with information matrix $\Lambda = J^\top W J$ (the weighted Jacobian of the rows), and the standard deviation of an unknown is the square root of the corresponding diagonal entry of Λ^{-1}. Treating the state as Gaussian around the MAP is the *Laplace approximation*. When several models compete — a leak here, a leak there, no leak — each is compared by its *evidence*, the probability of the readings under the model once its extra freedom is paid for: in the Laplace approximation

$$\log Z = -C(z^*) + \tfrac{d}{2}\log 2\pi - \tfrac{1}{2}\log\det\Lambda + \sum_k \tfrac{1}{2}\log\frac{w_k}{2\pi},$$

with d the number of unknowns and the last sum over every row so that models of different size compare as densities. Differences of $\log Z$ are in *nats* (natural-log units of odds; 5 nats is odds of about 150 to 1), and the models' posterior probabilities are their evidences normalized to sum to one under equal prior odds.

3.3 The two dimensionings

The sizing rule sets each link's conductance to the heat it carries at the design point divided by the temperature difference the design allows across it. With a design rack power P, N_{rk} racks per room, N_{room} rooms, a design air rise ΔT_{rise} of 12 K and coil approaches ΔT_{app} of 3 K:

$$G_{\mathrm{rk}} = P/\Delta T_{\mathrm{rise}}, \tag{19}$$

$$G_{\mathrm{room}} = N_{\mathrm{rk}}\, P/\Delta T_{\mathrm{app}}, \tag{20}$$

$$G_{\mathrm{bldg}} = N_{\mathrm{room}}\, N_{\mathrm{rk}}\, P/\Delta T_{\mathrm{app}}. \tag{21}$$

The specific heat of air cancels in Eq. (19): a link that carries heat P across a temperature difference ΔT has conductance $P/\Delta T$ whatever the mass flow that achieves it. The rule is applied once per model, at that model's own design rack power and geometry; Table 4 lists the resulting rack conductances. The low-conductance condition is the same models with the conductances of Table 1, which carry no such relation to their design heat.

Sizing also moves the point at which a room becomes unstable. Leakage feeds back: a warmer rack draws more power, which warms it further. For a room of N_{rk} identical racks of leakage slope α the feedback gain reaches one at

$$\alpha^* = \frac{G_{\mathrm{rk}}\, G_{\mathrm{room}}}{N_{\mathrm{rk}}\, G_{\mathrm{rk}} + G_{\mathrm{room}}}. \tag{22}$$

Equation (22) assumes the node beyond the room link is held at a fixed temperature; a finite building link shared by several rooms lowers the slope at which runaway sets in, so (22) is an upper bound on it. The condition $G_{\mathrm{rk}} > \alpha$, which the sizing helper enforces before it returns, is only the necessary condition for a single rack; it does not imply $\alpha < \alpha^*$ for the room.

3.4 The temperature screen

Every model is run under a screen that records the temperature of every node of every state the model solves, whether that state is reported or not, and keeps the hottest and the coldest. Two models search over states they never report — a capacity search that pushes load past a rack limit to find where the limit binds, and a cross-check that re-solves a sweep already reported — and the screen records those solves separately, so that the hottest node of a reported state is not confused with the hottest node of a search. A model's screened temperatures are then held to a band, and Table 2 gives the classes of band the collection needs: the 320 K air limit for a model whose reported states are all operating states, and three classes for models whose reported states are hotter by construction. The band classes were assigned after the temperatures had been measured.

3.5 Re-derivation and survival

A model whose hottest node lay above the air limit under the low-conductance dimensioning is an affected model, and there are eighteen of them. For each one, the conclusion it reported is recorded as a sentence together with the numbers that carry it, each read from the report the low-conductance model produced at the precision that report printed. The same quantities are then read from the sized model, and three things are computed. A number is *equal at printed precision* when the two values agree once both are written with the digits the first was printed with. A quantity whose claim is carried by its side of a reference value — a speed-up against equal cost, a rank against first place, a feedback gain against one, a projected date against the true one — has a *direction*, kept when both values lie on the same side of that reference and reversed otherwise. The *magnitude ratio* is the sized value over the low-conductance value.

Two judgments are recorded by hand for each model, each with its reason: whether the reported conclusion was wrong independently of the dimensioning, and whether it still holds under the sized dimensioning. The category is then the first of these rules that matches, in this order: *wrong independent of sizing*, when the conclusion was already wrong under the low-conductance dimensioning, which is therefore never counted as an effect of sizing; *reversed by sizing*, when a direction reversed or the conclusion no longer holds; *framing changed*, when every headline number is equal at printed precision and only the region in which the conclusion is valid has moved; *magnitude changed*, when the direction is kept and some headline number differs; and *unchanged* otherwise. The categories are author judgments on which conclusion each model's report leads with, and Section 5 states this as a threat.

3.6 The mechanism experiment

Why diagnostic evidence shrinks when links are sized is answered on one model of the collection, the three-room, four-rack building of Figure 1, built twice with the same racks, sensors, noise levels and priors and once per dimensioning (10, 20 and 100 W/K against 45.8, 733 and 2200 W/K). Three faults are planted in turn, at a rack, a room and the building, with the magnitudes 250 W, 1,000 W, 2,500 W respectively. Observations are noise-free: every sensor reads the fault-free-plus-fault solve exactly, so that the experiment measures the model's ability to separate faults and not a noise realization. Each fault is estimated twice: with only the true fault latent, and with every fault latent, as a diagnosis runs it. Two measures of evidence are reported, the true magnitude and the estimate in units of the posterior standard deviation:

$$z_F = F/\mathrm{sd}, \tag{23}$$

$$z_\mu = |\mu|/\mathrm{sd}. \tag{24}$$

Three closed forms are computed alongside them, all whitened by each sensor's own noise so that they are dimensionless per watt. Writing $\tilde{s}$ for the whitened sensitivity of the sensors to the true fault, $\tilde{S}$ for the whitened sensitivities to the competing faults and Λ_c for the diagonal matrix of the competing faults' prior precisions, the transfer norm, the projection that removes what the competing faults can explain, and that projection with the prior terms restored are

$$\|\tilde{s}\| = (\tilde{s}^\top \tilde{s})^{1/2}, \tag{25}$$

$$q_0^2 = \tilde{s}^\top \tilde{s} - \tilde{s}^\top \tilde{S}(\tilde{S}^\top \tilde{S})^{-1} \tilde{S}^\top \tilde{s}, \tag{26}$$

$$q^2 = \tilde{s}^\top \tilde{s} - \tilde{s}^\top \tilde{S}(\tilde{S}^\top \tilde{S} + \Lambda_c)^{-1} \tilde{S}^\top \tilde{s} + 1/s^2. \tag{27}$$

The models are linear in the heat terms, so the Laplace approximation is exact here and the evidence measure follows the projected sensitivity: $z_F = F\,q$ with every fault latent, and the same with $\tilde{S}$ empty when only the true fault is. The experiment measures both identities. The largest relative miss it finds, over both dimensionings and all three levels, is 0.000149 in the single-latent case, which comes from the tolerance with which the physical rows are weighted, and below 10^{-5} with every fault latent; against the tolerance of 10^{-3} at which the experiment fails, that is a margin of 6.7. The stored results keep the power-of-ten bounds rather than the measured digits, which move between floating-point kernels. Fault magnitudes, sensor noise and priors are held fixed across the two dimensionings, so every ratio the experiment reports is an effect of the conductances alone.

3.7 The timing protocol

Three models of the collection report a cost comparison: keeping a model alive against rebuilding it for each what-if, a structure-aware sparse solve against a dense one, and a rank-one update of a posterior against a full re-solve. Each defines a bound as an extreme statistic — the smallest ratio over the sizes measured, or the fastest of one workflow against the slowest of the other — and publishes the integer part of it, so that the claim is a floor. Under the sized dimensioning the whole protocol was repeated five times per model, each repeat starting only after a check found no other computing process on the machine and its processor load, averaged over ten seconds, below a fixed ceiling, with a single-threaded linear-algebra library. The published bound is the smallest of the five repeats, and the range over the repeats is reported with it. The low-conductance factors these are compared against were each a single measurement on a machine that was not checked, so they bound nothing.

3.8 Metrics

Four measures answer the three questions. *Plausibility* is the count of models whose hottest solved node is at or below the 320 K air limit, out of 45; the reference is the whole collection, since an air-cooled rack inlet above that limit is outside the operating range such facilities are run in (Ebrahimi et al. 2014). *Survival* is the category of each affected conclusion, counted over the eighteen affected models; the reference is *unchanged*, what a conclusion that does not depend on the dimensioning would give, and the magnitude ratio measures how far a surviving conclusion moved, with one meaning it did not move. *Evidence* is z_F of Eq. (23), the fault magnitude in units of the posterior standard deviation it is estimated with, and the quantity of interest is its sized-over-low-conductance ratio, to be read against the ratio of the link conductance, which is what a single-link argument predicts. *Cost* is the floor of each comparison's ratio, to be read against one, equal cost. Counts are given as *k of n* and the comparison of dimensionings is a sensitivity analysis: it says how the conclusions depend on one modeling choice, and it is not a statement of uncertainty about them.

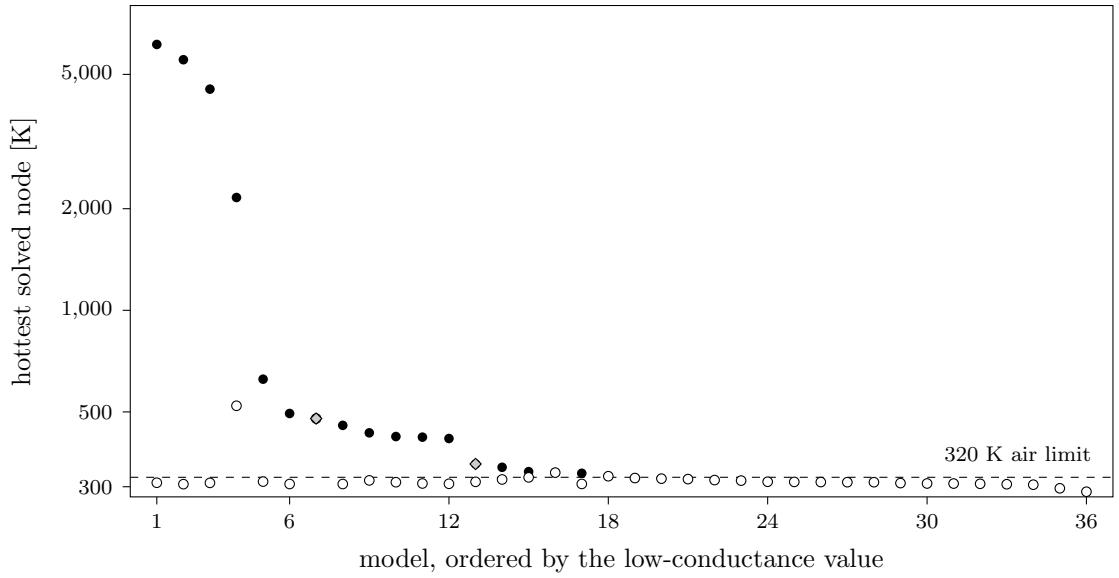


Figure 3: The hottest solved node of each of the thirty-six models that solve a thermal network, under both dimensionings, on a logarithmic vertical axis. Models are ordered by their low-conductance value, so the left of the figure holds the models that solved furthest from a physical temperature. Filled circles are the low-conductance dimensioning, open circles the sized one, and the dashed line is the 320 K air limit. Gray diamonds mark the two models whose hottest solve is made inside a search over states that the model does not report. The hottest low-conductance value is 6,134 K (`graph_structure_scaling`), and the coldest is −21,156.8 K (`incremental_inference`), a network past thermal runaway. Under the sized dimensioning every model whose band is the air limit sits at or below 319.9 K.

4 Results

4.1 Plausibility of the solved temperatures

Figure 3 plots the hottest solved node of each of the thirty-six models that solve a thermal network, once per dimensioning, and Table 2 gives the band classes with their counts.

Under the low-conductance dimensioning 18 of 45 models keep every solved node at or below the 320 K air limit and 18 do not, which are the affected models of Section 3.5. The excursions are large rather than marginal: the hottest solved node is 6,134 K, and three models solve at a coldest node below 0 K, down to −21,156.8 K, which is what a network past its runaway point gives — the steady state it converges to is the unphysical branch of Eq. (22). Under the sized dimensioning 32 of 45 models are at or below the air limit, against 18 before. The remainder is accounted for: 9 models solve no thermal network and have no screened temperature, 2 report a fault state whose air is hotter than the limit on purpose, 1 reports and marks states beyond a device limit, and 1 has the approach to runaway as its subject, so its band is a mathematical bound rather than a survivable temperature. Of the 18 affected models, 14 now sit at or below the air limit and the other 4 are exactly those hot-by-design cases. No single rule did this: thirteen of the eighteen models were re-dimensioned by the rule of Eqs. (19)–(21), one had its temperature response rescaled, one gained a device limit that marks its hot sweep points infeasible, and three were already reporting hot states by design and received a band derived from their own scenario.

Table 2: The band every model's screened temperatures are held to, by class, with counts over the forty-five models of the collection. The band of a model in the first class is the 320 K air limit; the other three classes hold models whose reported states are hotter by construction, and their members are named. The classes were assigned after the temperatures had been measured, which Section 5 states as a threat.

class	models	what the band means	models
air limit	32	every solved node of a reported state must sit at or below the air limit	—
not screened	9	the model solves no thermal network, so no temperature is screened	—
reported fault state	2	the model reports a fault state whose air is hotter than the limit by design	`airflow_redistribution`, `device_placement`
reported infeasible state	1	the model reports and plots states beyond a device limit, marked infeasible	`fan_psu_optimization`
runaway bound	1	the model's subject is the approach to thermal runaway, and its band is a bound, not a survivable temperature	`capacity_planning`

4.2 Survival of the conclusions

Table 3 lists the eighteen affected conclusions with their headline quantity under both dimensionings, their direction, their magnitude ratio and their category.

3 of 18 conclusions are unchanged and 7 keep their direction with a different magnitude, so 10 of 18 survive in the sense that the reader of the report would draw the same conclusion. 1 changes only the region in which it is valid, 2 reverse, and 5 were wrong independently of the dimensioning and are therefore not counted as effects of sizing at all. The magnitudes that moved moved far: the magnitude ratios of the surviving conclusions span 0.0319 to 2.52, so a claim can keep its direction and lose a factor of thirty.

The two reversals are the most consequential findings, and both are diagnostic. `sensor_placement_fault_isolation` reported that two sensors chosen by expected information gain raise the true explanation from rank 3 (21.2%) to rank 1 (74.1%), against 21.1% and 27.6% for the highest-load and uniform-row placements. Its leading number, 74.1% as reported, is 0.236 under the sized dimensioning, a ratio of 0.319. Under the sized dimensioning after the two sensors the true explanation ranks second, behind a joint row-and-rack explanation that calls for the same rack repair. `degradation_forecasting` reported that on day 90 the projected threshold crossing is day 137 (band 136 to 138) against a true crossing on day 123; the band narrows from 18.4 to 4.7 to 1.9 days, and the projected date reads as an upper bound on the crossing. Its leading number, 1.9 as reported, is 14.5 under the sized dimensioning, a ratio of 7.63. Under the sized dimensioning the realized day-90 projection falls before the true crossing and its band contains it, so the projected date no longer bounds the crossing from above; the noise-free projection is still late.

Five conclusions do not depend on the dimensioning at all: they were wrong under the low-conductance models as well, and re-deriving every number is what found them. `containment_retrofit` reported that sealing the aisles cools every rack by about 14.0 K, trims 252 W (4.1 %) of IT power and cuts the feedback gain 3.1-fold; most of the benefit arrives with the first round of sealing. Its leading number, 14.0 as reported, is 4.1 under the sized dimensioning, a ratio of 0.29. That conclusion was already wrong: the saving curve is steepest near perfect sealing, not at the leaky end, so a first round of sealing captures a small share of the benefit at either dimensioning. `graph_structure_scaling` reported that the structure-aware sparse solve is 11 times faster than the dense solve at 1,024 racks with the same answer, because the interface between rooms and

Table 3: Survival of the eighteen conclusions that the two dimensionings can disagree about, ordered by category. The headline quantity is each model's leading number, given as the value the low-conductance model reported and the value the sized model gives, both at the precision the first was printed with; the ratio is the second over the first. Direction says whether the quantity stayed on the same side of the reference that carries the claim (Section 3.5); the three models at the foot report no conclusion that sizing can move, so they have none. Categories are assigned by the rules of Section 3.5 in their order of precedence, and the counts are 5 wrong independent of sizing, 2 reversed by sizing, 1 framing changed, 7 magnitude changed, 3 unchanged.

model	headline quantity	direction	ratio	category
`containment_retrofit`	14.0 → 4.1	kept	0.29	wrong independent of sizing
`graph_structure_scaling`	11 → 20	kept	1.82	wrong independent of sizing
`incremental_update`	4.4 → 1.0	kept	0.227	wrong independent of sizing
`capacity_planning`	0.88 → 0.64	kept	0.73	wrong independent of sizing
`fan_psu_optimization`	121 → 121	kept	1	wrong independent of sizing
`sensor_placement_fault_isolation`	74.1% → 0.236	reversed	0.319	reversed by sizing
`degradation_forecasting`	1.9 → 14.5	reversed	7.63	reversed by sizing
`when_laplace_lies`	0.14 → 0.14	kept	1.01	framing changed
`active_sensing`	693 → 743	kept	1.07	magnitude changed
`building_fault_isolation`	466.8 → 100.1	kept	0.214	magnitude changed
`incremental_inference`	37 → 9	kept	0.243	magnitude changed
`sensor_placement_optimization`	5.2 → 13.1	kept	2.52	magnitude changed
`thermal_fault_localization`	559.4 → 102.4	kept	0.183	magnitude changed
`bimodal_fault_isolation`	1.40 → 1.27	kept	0.908	magnitude changed
`containment_effectiveness`	97.2 → 3.1	kept	0.0319	magnitude changed
`stranded_capacity`	—	—	—	unchanged
`airflow_redistribution`	—	—	—	unchanged
`device_placement`	—	—	—	unchanged

trunk stays small (1033 shared quantities). Its leading number, 11 as reported, is 20 under the sized dimensioning, a ratio of 1.82. That conclusion was already wrong: the interface count is fixed by the topology and identical under both dimensionings; it is about a third of the unknowns, so it never stayed small. `incremental_update` reported that keeping the model alive gives the same answer as rebuilding it for every what-if and is 4.4 to 6.2 times cheaper per event across sizes (5.1 times at 1,024 racks). Its leading number, 4.4 as reported, is 1.0 under the sized dimensioning, a ratio of 0.227. That conclusion was already wrong: each room's cooling link ended at a reservoir of its own, so no heat reached the building node and it printed the chilled-water temperature at any conductance. `capacity_planning` reported that runaway sets in at a leakage slope of 1.67 W/K; the feedback gain is 0.05 today and 0.88 in the degraded room, and the room-temperature curves stop at the tipping point. Its leading number, 0.88 as reported, is 0.64 under the sized dimensioning, a ratio of 0.73. That conclusion was already wrong: the published room-temperature sweeps plotted converged solutions past the tipping point, where no steady state exists, while the text said the curves stop. `fan_psu_optimization` reported that an interior optimum at fan speed 0.30 saves 121 W (22 %) against full speed; the best speed rises with workload (0.28, 0.30, 0.33) and follows each supply's efficiency sweet spot. Its leading number, 121 as reported, is 121 under the sized dimensioning, a ratio of 1. That conclusion was already wrong: the three supply classes, peaking at different loads, share one best speed in the published sweep itself, so the optimum never followed the supply's sweet spot.

One conclusion keeps every headline number and moves only the region in which it holds. `when_laplace_lies` reported that the Laplace self-check is valid below a sensor accuracy of 3 K

and flagged at and above it, with curvature scores from 0.05 to 2.90; near the runaway cliff the band is valid at 0.0026 W/K. Its leading number, 0.14 as reported, is 0.14 under the sized dimensioning, a ratio of 1.01. Under the sized dimensioning the response was rescaled with the sensor accuracy and physics tolerance, so every statement about the estimated slope is unchanged; the boundary is stated at the rescaled accuracy. The published claim's statement about the particle filter is withdrawn rather than re-scored: it described an engine that has since been corrected, which is not a consequence of the dimensioning.

The seven conclusions that kept their direction with a different magnitude are the diagnosis, placement and containment models. Their headline numbers shrink or grow by the ratios of Table 3: the diagnosis models' evidence falls by factors of 5 to 29, which Section 4.3 explains; the containment model's leakage penalty falls by about 31; and the sensor-placement model's advantage over the alternatives grows, from 5.2 % to 13.1 % less remaining variance than a random layout, because with sized links a rack sensor sees its own rack rather than the whole room. Three conclusions are unchanged: their models carry no conductance the rule applies to, they report the hot states of their own scenarios, and their hottest solved node is the same under both dimensionings.

4.3 Why the evidence shrinks

Figure 4 shows the controlled experiment of Section 3.6: per level of the fault, the sized-over-low-conductance ratio of the fault's own link conductance, of the whitened transfer norm, of the projected form q_0 and of the evidence measure z_F with every fault latent, together with the ratios of the top z that the three-room diagnosis reports under measurement noise.

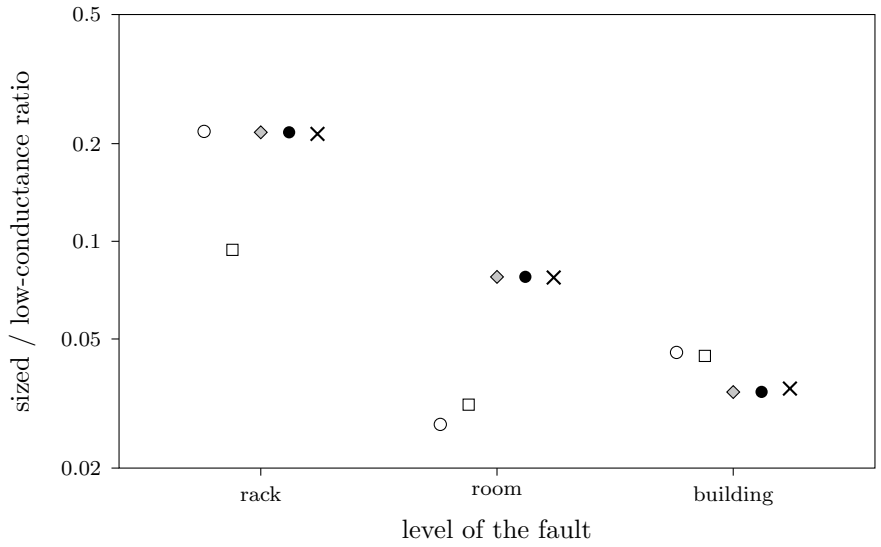
A single-link argument predicts that sizing shrinks a fault's signature by the ratio of its link conductance, because a fault of ΔQ across a link of conductance G raises a temperature by $\Delta Q/G$. That prediction holds at the rack level and fails at the other two. With only the true fault latent, the evidence ratio follows the whitened transfer norm exactly, as Eq. (25) says it must: 0.09413 against 0.09411 at the rack level, 0.03137 against 0.03137 at the room level and 0.04435 against 0.04435 at the building level. With every fault latent, which is how a diagnosis runs, the ratios are 0.2167, 0.07773 and 0.03434 instead, and they follow the projected form q of Eq. (27) to within 10^{-5} relative. The evidence for a fault is carried not by how strongly the sensors respond to it but by the part of that response which the competing faults cannot reproduce.

At the rack level that part is the contrast between one rack and its neighbors, which is set by the rack link alone, so the ratio 0.2167 comes out close to the link ratio 0.2182: the local-link rule is an approximation good to 0.7 % here, not an identity. At the room level the measured ratio 0.07773 is 2.5 times the transfer norm's 0.03137: sizing costs a room fault less evidence than its own response suggests, because the room's signature becomes easier to distinguish from the racks'. At the building level the measured ratio 0.03434 is below both the link ratio 0.04546 and the transfer norm 0.04435. The three crosses in Figure 4 are the realized top-z ratios of the diagnosis model itself, computed with measurement noise and with all the model's faults competing: 0.2143, 0.07734 and 0.03517. They agree with the noise-free experiment to within 2 %, so the shrinkage the diagnosis models report is the projection effect and not a noise realization.

4.4 Cost of the solves

Figure 5 plots the four cost comparisons of the collection: each published bound under the sized dimensioning with its range over the five repeats, against the factor the same comparison carried under the low-conductance dimensioning.

Two of the four bounds fell and two rose. For `incremental_update`, comparing the per-what-if



open circle: local link ratio; open square: whitened transfer norm; gray diamond: projected form q_0; filled circle: measured evidence ratio, all faults latent; cross: realized top z under measurement noise

Figure 4: How much a fault's evidence shrinks when the conductances are sized, per level of the fault, on a logarithmic vertical axis. Each ratio is the sized value over the low-conductance value of the same quantity in the same noise-free experiment. Open circles are the ratio of the fault's own link conductance, open squares the whitened transfer norm $\|\tilde{s}\|$ of Eq. (25), gray diamonds the projected form q_0 of Eq. (26) and filled circles the evidence ratio measured with every fault latent, as a diagnosis runs it. Crosses are the ratios of the top z that the three-room diagnosis reports with measurement noise, at the same three levels, and they agree with the noise-free experiment. At the rack level the measured ratio 0.2167 follows the local link ratio 0.2182; at the room and building levels it follows neither the link nor the transfer norm (0.07773 against 0.03137, 0.03434 against 0.04435) and follows the projected form.

cost of keeping a model alive against rebuilding it, smallest over the sizes measured, the published floor is 1× (1–3× over the repeats) against 4.4× under the low-conductance dimensioning. For `graph_structure_scaling`, comparing a structure-aware sparse solve against a dense one at the largest size measured, the published floor is 20× (20–23× over the repeats) against 11× under the low-conductance dimensioning. For `incremental_inference`, comparing a rank-one posterior update against a full re-solve, on every streamed reading, the published floor is 9× (9–25× over the repeats) against 37× under the low-conductance dimensioning. For `incremental_inference`, comparing a rank-one posterior update against a full re-solve, at every model size, the published floor is 13× (13–21× over the repeats) against 7× under the low-conductance dimensioning.

The spread over the repeats is itself the finding: a bound defined as a minimum over sizes or a fastest against a slowest is an extreme statistic, and on a machine checked to be quiet it still moved by up to a factor of 3.0 between repeats minutes apart. The bounds are therefore published as the minimum of five repeats and read as floors, not as measurements of the advantage. They are not comparable to the low-conductance factors as measurements: those were single runs on an unchecked machine, and both the model and the machine differ.

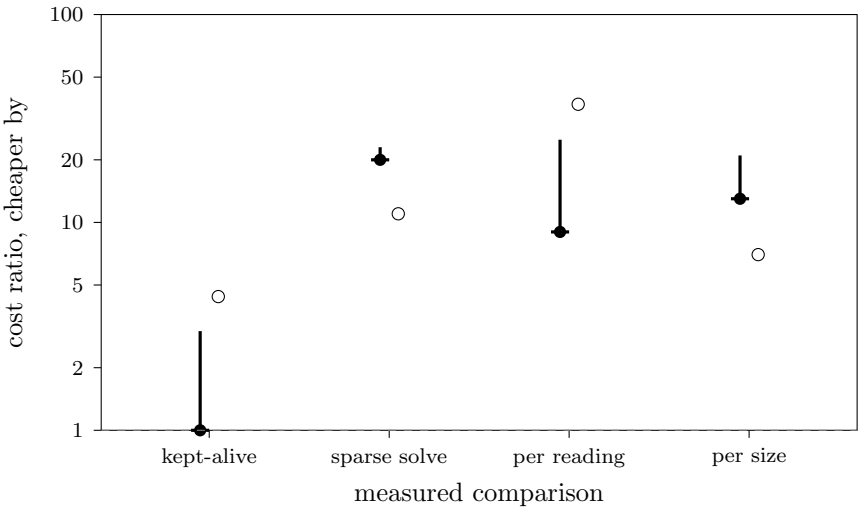


Figure 5: The four measured cost comparisons of the collection, on a logarithmic vertical axis. Each filled circle is the published bound under the sized dimensioning, the minimum over five repeats of the whole protocol, and the bar above it reaches that bound's maximum over the repeats. Open circles are the factors the same comparisons carried under the low-conductance dimensioning, each a single measurement on a machine that was not checked to be quiet, so they bound nothing; the dashed line is equal cost. The first two comparisons are of a kept-alive model against a rebuilt one and of a structure-aware sparse solve against a dense one; the last two are of a rank-one posterior update against a full re-solve, per reading and per model size.

5 Discussion

The comparison establishes that one modeling choice decided the temperatures of 18 of 45 models and the magnitude of most of their conclusions, while reversing 2 of them. It also shows what such a change does not do: 10 conclusions survive with their direction, and the models' structural claims — which quantities are identifiable, which layout of sensors is best, how the cost of a solve scales — are the ones that survive most cleanly. The mechanism experiment explains the pattern: evidence follows the projected sensitivity of Eq. (27), so a conclusion that rests on a contrast between neighbors moves with the local link, and one that rests on a level's total response moves by a factor that only the projection predicts. This also says what to check when a compact thermal model is built for diagnosis: not that the temperatures look reasonable, but that each link's conductance carries its design heat, because the diagnosis' confidence is a function of those conductances.

Several threats limit how far this can be read.

- *The low-conductance values are not always unambiguous.* Each model's reported numbers were read from the report that model produced, at the precision it printed, and a few of those numbers also appeared elsewhere in that collection's own summaries with different values. Only the report-derived values are used here, and a value that could not be found in the report was left out of the comparison altogether.

- *The categories are author judgments.* The direction and the magnitude ratio are computed, but which quantity carries a conclusion, and whether a conclusion still holds, are deci-

sions about what each report leads with. Two of them are worth naming. `degradation_forecasting` is counted as a reversal because its realized projection moved to the other side of the true crossing date, although the noise-free projection remains on the reported side and the realized error is in the conservative direction. `sensor_placement_fault_isolation` is counted as a reversal at the level of the exact explanation, while its action-level conclusion, which repair the crew should perform, still holds. Recording the flag and its reason per model, and computing the category from them, makes these judgments visible but does not remove them.

- *Band classes were assigned after measurement.* The three classes other than the air limit were introduced once the screened temperatures were known. They are justified by what each model reports — a fault state, an infeasible state, a runaway bound — but a class introduced after the measurement cannot also be evidence that the measurement passed.

- *The mechanism analysis was designed after the ratios were known.* The evidence ratios of the diagnosis models were known before the closed forms of Eqs. (25)–(27) were written down. The identities are exact for a linear-Gaussian model and are checked numerically, so their correctness does not depend on the order of work, but no pre-registered hypothesis is claimed and the local-link rule at the rack level is an approximation that was found, not predicted.

- *One equivalence is judged on a yardstick chosen after the fact.* The model that compares a rank-one posterior update with a full re-solve judges their agreement against each estimate's posterior standard deviation. Under the sized dimensioning the agreement holds on that yardstick and fails on the one the model reported earlier, the size of the update itself. The two yardsticks differ by orders of magnitude on a sized network, and the choice between them is a judgment this report inherits rather than settles.

- *Evidence is measured on realized draws in the models and noise-free in the experiment.* Equations (23) and (24) are two different measures, the fault magnitude and the estimate, in units of the same posterior standard deviation; the diagnosis models report the realized estimate under one noise draw, while the experiment removes noise to isolate the projection. They agree here to within 2 %, but a single draw can differ from the noise-free value by more than that, and no distribution over draws is computed.

- *The costs come from one machine.* All four bounds were measured on one workstation with a single-threaded linear-algebra library, and the ranges of Figure 5 show how much they move between repeats on it. They bound the advantage of one implementation against another on that machine and do not predict it elsewhere.

- *The collection is synthetic and shares its physics.* Every model generates its own data from the same family of relations, so the comparison varies a conductance inside a modeling family and cannot tell whether that family is right for a real room. Errors common to the family — the symmetric link of Figure 2, perfectly mixed room air, a leakage law linear in temperature — move both dimensionings alike and are invisible here.

Two smaller points bear on reproducibility. The agreement of the closed forms with the measured evidence, and several other quantities in this collection, sit at the level of floating-point round-off, where results depend on which kernel of the linear-algebra library runs (Harris et al. 2020); they are reported here as power-of-ten bounds for that reason. And the sized dimensioning makes the networks stiffer: a conductance that grows by a factor of a thousand makes the fit's

normal equations correspondingly worse conditioned, which is the reason the equivalence yardstick above had to be reconsidered at all.

6 Conclusion

This report asked whether the solved temperatures of a collection of compact data-center thermal models are physical, whether their conclusions survive a change in the conductances they carry, and why their diagnostic evidence changes as it does. Under a low-conductance dimensioning 18 of 45 models keep every solved node at or below the 320 K air limit, with the hottest solving at 6,134 K; sizing every link for the heat it carries, at a 12 K rack air rise and 3 K coil approaches, brings that count to 32 of 45, and the remainder are the 9 models that solve no network and the 4 whose hot states are reported by design. Of the 18 conclusions the change can move, 3 are unchanged, 7 keep their direction with magnitudes that moved by factors of up to 31, 1 keeps every number and moves its validity region, 2 reverse, and 5 were wrong independently of the dimensioning. The evidence for a fault shrinks by 0.2167, 0.07773 and 0.03434 at the rack, room and building levels, and those ratios follow the part of the fault's sensitivity that competing faults cannot explain rather than its own link conductance or its total response, an identity the experiment reproduces to within 10^{-5} relative with every fault latent. The cost comparisons, re-measured on a quiet machine with five repeats, give floors of 1×, 20×, 9×, 13×, which move by up to a factor of 3.0 between repeats and are read as floors rather than measurements. The next steps are a comparison against a flow-resolved model, which would test the symmetric-link assumption the whole family shares, and a distribution over noise draws for the evidence measures, which this report computes at a single draw.

A Reproduction

```
cd docs/terra_graph_engine/case_studies/cross_level/thermal_network_dimensioning
PYTHONPATH=../../../../../src python run.py
PYTHONPATH=../../../../../src python mechanism_experiment.py
PYTHONPATH=../../../../../src python make_report.py --pdf
cd ../../../../..
python -m pytest -q tests/plugins/datacenter/test_thermal_network_dimensioning_case_study.py
```

`run.py` reads the committed files listed in Appendix B, runs the mechanism experiment of Section 3.6 and writes `results.json`; with `--out DIR` it writes into another directory instead. `mechanism_experiment.py` runs that experiment alone and prints its record. `make_report.py` reads only `results.json` and writes `report.tex`. Every model of the collection has its own entry point and its own test, and the screen that measures the temperatures of Figure 3 is a separate command that runs each model in a child process. The experiment is deterministic: it plants each fault in turn and reads the sensors of the noise-free solve, with no random draw anywhere, and `results.json` is identical byte for byte under the two floating-point kernels of the linear-algebra library available on the machine used. Numbers in `results.json` carry 6 significant digits, except the quantities at the level of round-off, which are stored as the bounds this report prints. The test in the last line regenerates the record, compares it with the committed one and rebuilds this report.

B Data and provenance

No measured data enter this report. Its inputs are committed files: the band registry of the collection, the temperature screen of all 45 models under the sized dimensioning, the same screen under the low-conductance dimensioning, the eighteen affected models' own result files, and the three timing files. The low-conductance temperatures of Figure 3 and Table 4 are that screen as it was recorded at commit `84344781`, before any conductance changed. Every low-conductance number in Table 3 and in Section 4.2 was read from the report the model in question produced at commit `5be377ed`, at the precision that report printed it with, and each is stored beside the value it is compared against, with the key that names it in that model's own result file; values that could not be found in the report itself are not used. Conductances under the sized dimensioning follow the rule of Eqs. (19)–(21). Table 4 lists every model of the collection. The collection leaves three problems open: a comparison of one of these networks against a flow-resolved model of the same room, a distribution of the evidence measures over noise draws rather than the single draw each diagnosis model reports, and a re-measurement of the four cost comparisons on a second machine.

C Notation

Table 5 lists the symbols of the report.

References

[Barroso et al. 2019] Barroso, L. A., Hölzle, U., Ranganathan, P. (2019). The Datacenter as a Computer: Designing Warehouse-Scale Machines, 3rd edition. Synthesis Lectures on Computer Architecture, Springer. doi:10.1007/978-3-031-01761-2.

[Ebrahimi et al. 2014] Ebrahimi, K., Jones, G. F., Fleischer, A. S. (2014). A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. Renewable and Sustainable Energy Reviews 31, 622–638. doi:10.1016/j.rser.2013.12.007.

[Fulpagare and Bhargav 2015] Fulpagare, Y., Bhargav, A. (2015). Advances in data center thermal management. Renewable and Sustainable Energy Reviews 43, 981–996. doi:10.1016/j.rser.2014.11.056.

[Harris et al. 2020] Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. Nature 585(7825), 357–362. doi:10.1038/s41586-020-2649-2.

[Krause et al. 2008] Krause, A., Singh, A., Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies. Journal of Machine Learning Research 9, 235–284. https://www.jmlr.org/papers/v9/krause08a.html.

[Kschischang et al. 2001] Kschischang, F. R., Frey, B. J., Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory 47(2), 498–519. doi:10.1109/18.910572.

[Masanet et al. 2020] Masanet, E., Shehabi, A., Lei, N., Smith, S., Koomey, J. (2020). Recalibrating global data center energy-use estimates. Science 367(6481), 984–986. doi:10.1126/science.aba3758.

Table 4: Every model of the collection, one row each. The two temperature columns are the hottest node the model solves under the low-conductance and the sized dimensioning, at the screen's precision of 0.1 K; the band column is the upper bound its class allows. The sizing column says how a model whose low-conductance temperature lay above the 320 K air limit was brought under it, and the last column gives its rack-link conductance before and after sizing where the model has one. The class column uses the short forms *fault state*, *infeasible* and *runaway* for the three classes of Table 2 other than the air limit, and the sizing column says whether a model above the limit was brought under it by the sizing rule (*rule*), by a rescaled temperature response (*rescaled*), by a device limit (*limit*), or not at all because its hot states are reported by design (*by design*). A dash means the quantity does not exist for that model: it solves no thermal network, its temperatures were already physical, or its links carry no design heat.

level	model	low-G [K]	sized [K]	band [K]	class	sizing	G_{rk} [W/K]
L1	active_sensing	332.7	319.9	320.0	air limit	rule	50 / 54.2
L1	bad_sensor_trust	310.1	310.1	320.0	air limit	—	—
L1	building_codesign	—	—	—	not screened	—	—
L1	building_fault_isolation	494.5	305.7	320.0	air limit	rule	10 / 45.8
L1	containment_retrofit	342.5	315.3	320.0	air limit	rule	15 / 25.0
L1	cyclic_facility_hotspot	304.3	304.3	320.0	air limit	—	—
L1	diagnosis_multi_crah	309.5	309.5	320.0	air limit	—	—
L1	graph_structure_scaling	6,134.4	308.3	320.0	air limit	rule	10 / 45.1
L1	hidden_hotspot_decision	306.7	306.7	320.0	air limit	—	—
L1	incremental_inference	4,522.2	308.1	320.0	air limit	rule	10 / 25.0
L1	incremental_update	5,526.0	305.3	320.0	air limit	rule	10 / 53.0
L1	rack_hotspot_weather	305.2	305.2	320.0	air limit	—	—
L1	real_weather_free_cooling	309.8	309.8	320.0	air limit	—	—
L1	scaling_to_a_building	310.5	310.5	320.0	air limit	—	—
L1	sensor_placement_fault_isolation	420.9	306.8	320.0	air limit	rule	10 / 52.5
L1	sensor_placement_optimization	456.0	305.7	320.0	air limit	rule	10 / 45.8
L1	stranded_capacity	350.4	310.2	320.0	air limit	by design	—
L1	thermal_fault_localization	417.0	306.1	320.0	air limit	rule	10 / 54.2
L1	weather_driven_pue	318.3	318.3	320.0	air limit	—	—
L2	airflow_redistribution	330.4	330.4	330.9	fault state	by design	—
L2	bimodal_fault_isolation	422.5	309.5	320.0	air limit	rule	10 / 45.8
L2	capacity_planning	2,158.8	521.3	521.8	runaway	rule	10 / 2.5
L2	cascade_control	309.5	309.5	320.0	air limit	—	—
L2	chilled_water_setpoint_envelope	312.9	312.9	320.0	air limit	—	—
L2	cop_drift	—	—	—	not screened	—	—
L2	device_placement	322.3	322.3	322.8	fault state	by design	—
L2	latent_blind_spot	290.0	290.0	320.0	air limit	—	—
L2	mpc_supervisory_control	—	—	—	not screened	—	—
L2	nonlinear_cliff	—	—	—	not screened	—	—
L2	setpoint_relaxation	307.5	307.5	320.0	air limit	—	—
L2	topology_detective	296.8	296.8	320.0	air limit	—	—
L2	uncertainty_aware_control_recommendation	—	—	—	not screened	—	—
L2	when_laplace_lies	433.5	313.4	320.0	air limit	rescaled	—
L3	containment_degradation_advisory	306.6	306.6	320.0	air limit	—	—
L3	containment_effectiveness	624.9	310.8	320.0	air limit	rule	10 / 25.0
L3	escalating_heatwave	316.1	316.1	320.0	air limit	—	—
L3	four_resolutions	306.1	306.1	320.0	air limit	—	—
L3	identifiability_and_probe_design	317.1	317.1	320.0	air limit	—	—
L3	multizone_control_crosscheck	—	—	—	not screened	—	—
L4	chance_constrained_cooling	—	—	—	not screened	—	—
L4	chance_constrained_hard	—	—	—	not screened	—	—
L4	degradation_forecasting	328.9	305.8	320.0	air limit	rule	10 / 25.0
L4	online_codesign	—	—	—	not screened	—	—
L4	the_rogue_rack	313.9	313.9	320.0	air limit	—	—
L5	fan_psu_optimization	477.8	477.8	478.3	infeasible	limit	—

Table 5: Symbols used in the report, with units.

symbol	meaning	unit
T_{rk}, T_{room}, T_{bldg}, T_{chw}	temperature of a rack, room air, the building loop, chilled water	K
P, N_{rk}, N_{room}	rack power at the design point, racks per room, rooms	W, —, —
G_{rk}, G_{room}, G_{bldg}	conductance of a rack, room and building link	W/K
G_{supply}, G_{return}	conductance of a row's supply and return legs	W/K
ΔT_{rise}, ΔT_{app}	design air rise across a rack, design coil approach	K
Q, f, s	heat flow on a link, unknown heat term, prior standard deviation of a heat term	W, W, W
α, α^*	leakage slope of a rack, leakage slope at which a room reaches gain one	W/K
y, σ, ε	sensor reading, sensor noise, the noise realized in a reading	K
F, μ, sd	true magnitude of a fault, its posterior mean, its posterior standard deviation	W
z_F, z_μ	evidence measures F/sd and $\lvert\mu\rvert$/sd	—
$\tilde{s}$, $\tilde{S}$, Λ_c	whitened sensitivity of the true fault, of the competing faults, their prior precisions	1/W, 1/W, 1/W^2
q, q_0	projected sensitivity with and without the prior terms	1/W
r_k, w_k, z, Λ	residual of a row, its weight, the state vector, the information matrix	—

[Oberkampf and Roy 2010] Oberkampf, W. L., Roy, C. J. (2010). Verification and Validation in Scientific Computing. Cambridge University Press. doi:10.1017/CBO9780511760396.

[Raue et al. 2009] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25(15), 1923–1929. doi:10.1093/bioinformatics/btp358.

[Saltelli et al. 2007] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. (2007). Global Sensitivity Analysis: The Primer. Wiley. doi:10.1002/9780470725184.

[Schweppe and Wildes 1970] Schweppe, F. C., Wildes, J. (1970). Power system static-state estimation, Part I: exact model. IEEE Transactions on Power Apparatus and Systems PAS-89(1), 120–125. doi:10.1109/TPAS.1970.292678.

[Tierney and Kadane 1986] Tierney, L., Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association 81(393), 82–86. doi:10.1080/01621459.1986.10478240.