Reliability of Gaussian uncertainty bands for a thermal-runaway parameter estimated from a single room-temperature reading

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

September 2026

Abstract

Estimates of physical parameters are routinely reported with a Gaussian uncertainty band obtained from the curvature of a fit, and the band is trusted without checking the Gaussian assumption. This report examines when such a band misstates the uncertainty for a parameter that controls thermal runaway: the leakage coefficient of a room whose equipment dissipates more heat as it warms, estimated from one room-temperature reading. It is an inference study. The coefficient is estimated as the maximum-a-posteriori value with a Laplace band, a curvature-and-importance self-check flags bands it judges unreliable, and every band is graded against the exact posterior, which a one-parameter problem admits by numerical integration. Over a sweep of the sensor accuracy from 0.05 K to 1 K at a moderate operating point, the self-check passes every band below 0.3 K and flags every band from 0.3 K on, its score rising from 0.052 to 2.90 against a gate of 0.25. The exact probability inside the nominal 95 % band is between 0.942 and 0.949 wherever the self-check passes and between 0.744 and 0.931 wherever it flags; the missing probability lies on the side away from runaway, where up to 25.6 % of it falls. Near runaway the band is narrow and passes, holding 0.947. The engine's particle filter, which samples the coefficient from the prior and integrates the room temperature out of every particle, agrees with the exact posterior at every point – within 0.19 of an exact standard deviation – so the band is graded against two references that share no machinery. The study costs 9 fits, 612 profile fits for the self-check and 72,000 particles.

1 Introduction

The power that electronic equipment dissipates rises with its temperature, largely through leakage current, and the rise feeds back: a warmer device leaks more, heats its surroundings more, and leaks more again. Where the cooling cannot remove the extra heat as fast as it appears, no steady temperature exists and the equipment runs away thermally. Vassighi and Sachdev (2006) describe the mechanism and its runaway condition for integrated circuits, and Zapater et al. (2015) show that the leakage–temperature coupling is large enough at server scale to change how cooling should be managed. A controller or an operator that keeps a room clear of runaway needs to know how strong the coupling is, and needs to know how well it knows it: an uncertainty band that is too narrow near the runaway condition invites operation closer to it than the evidence supports.

The most common uncertainty band for a fitted parameter is Gaussian, centered on the best fit and with a width taken from the curvature of the fit. Tierney and Kadane (1986) gave the accuracy of this Laplace approximation for posterior moments and marginal densities, and Rue et al. (2009) built fast approximate inference for a wide class of models on nested Laplace approximations. The approximation is exact when the model is linear in the parameter and degrades with

nonlinearity; Bates and Watts (1980) measured that nonlinearity by relative curvature. Whether a given approximation can be trusted is commonly checked by importance sampling, drawing from the approximation and reweighting by the exact density: Kong et al. (1994) introduced the effective sample size of such weights, and Yao et al. (2018) and Vehtari et al. (2024) turned the tail of the weights into a diagnostic. Particle methods, starting from the bootstrap filter of Gordon et al. (1993), avoid the Gaussian assumption altogether, at the cost of weight degeneracy when the likelihood is sharp relative to the proposal, which Snyder et al. (2008) quantified.

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 study uses it to write the steady heat balance of one room as a closure row, to fit the leakage coefficient from one reading with its Laplace band, to run the engine's self-check of that band, and to run the engine's bootstrap particle filter on the same model. Section 3 states the framework and each of these constructs precisely.

This report checks the Gaussian band and its self-check where the answer is known exactly. It is an inference study. It asks when the Laplace band on a parameter near a runaway condition misstates the uncertainty, and whether a curvature self-check detects it. The construction is a model with one parameter and one reading, whose exact posterior is obtained by numerical integration and used as the reference for the band, the self-check and the particle filter. Section 2 describes the room and the data, Section 3 the model, the self-check, the filter and the reference, Section 4 the results, and Sections 5 and 6 their limits and conclusions.

2 System and data

Figure 1 shows the system. A room holds equipment that dissipates a fixed heat $Q_0 = 1$ W plus a leakage heat proportional to how far the room is above the temperature of its cooling supply, $T_s = 300$ K; the proportionality is the leakage coefficient α in W/K. The room loses heat to the supply through a conductance $G = 2$ W/K. In steady state the heat removed equals the heat dissipated, so the room settles at a temperature that grows ever faster as α approaches G and has no steady value at or beyond $\alpha^* = G = 2$ W/K. The quantities are normalized: the heat and the conductance are chosen so that every room temperature the study meets lies between 300.9 and 313.8 K, and the question the report asks depends only on the ratio of the sensor accuracy to the size of the temperature response (Section 3.7).

One sensor reads the room temperature, and the task is to estimate α from that reading with an honest uncertainty band. The data are synthetic and deterministic. Two operating points are studied: a moderate one at $\alpha = 1.75$ W/K, where the room sits at 304.0 K, and one near runaway at $\alpha = 1.92$ W/K, where it sits at 312.5 K. At the moderate point the stated sensor accuracy is swept over 0.05, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1 K; near runaway it is 0.4 K, the accuracy at which the moderate point is already flagged. Table 1 lists every setting.

Figure 2 shows the model the estimate is computed on: two unknowns, the leakage coefficient and the room temperature, and three rows that tie them. The assumptions are these.

- The room is in steady state; the heat balance holds to the closure-row tolerance $\sigma_{\text{phys}} = 0.001$ K.

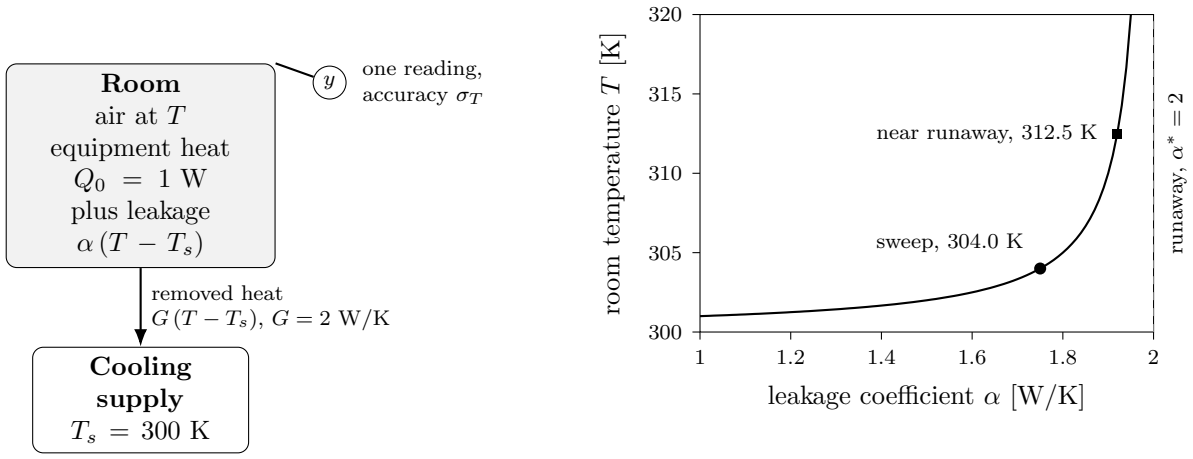


Figure 1: The system of this report. Left: a room whose equipment dissipates a fixed heat Q_0 plus a leakage heat that grows with the room's own temperature at the rate α; the room loses heat to a cooling supply at T_s through a conductance G, and one sensor reads the room temperature with stated accuracy σ_T. Right: the steady room temperature against α from Eq. (2). The curve steepens without bound toward the dashed line $\alpha^* = G$, where the steady state ceases to exist. The circle is the operating point of the sensor-accuracy sweep ($\alpha = 1.75$ W/K), the square the operating point near runaway ($\alpha = 1.92$ W/K); the slope at the square is 9.77 times the slope at the circle.

- The reading equals the model temperature at the true α. No noise is added; the stated accuracy σ_T sets only the weight of the reading, so every difference between the fit and the truth comes from the prior and from the shape of the model, not from a noise draw.
- The prior on α is Gaussian and broad compared with what the reading determines, and it is not truncated at α^*; the reading makes values beyond α^* negligible (Section 3.6).
- The room temperature carries no prior; it is known only through the closure row and the reading.
- Temperatures are in kelvin, heat in watts, α and G in W/K.

3 Methods

3.1 The room factor graph

The steady heat balance of the room and the temperature it implies are

$$G\,(T - T_s) = Q_0 + \alpha\,(T - T_s), \tag{1}$$

$$T(\alpha) = T_s + \frac{Q_0}{G - \alpha}. \tag{2}$$

$$\frac{dT}{d\alpha} = \frac{Q_0}{(G - \alpha)^2}. \tag{3}$$

The slope (3) is what makes the reading informative: a temperature uncertainty σ_T translates into an uncertainty of about σ_T divided by the slope on α, and the slope is 16 K per W/K at

Table 1: Every setting of the study. The reading handed to the fit is the model temperature at the true α, without added noise; σ_T is the accuracy the fit is told the sensor has, and so sets the weight of the reading.

quantity	value
supply temperature T_s	300 K
conductance to the supply G	2 W/K
equipment heat Q_0	1 W
runaway value $\alpha^* = G$	2 W/K
sweep: true α; prior mean and standard deviation	1.75; 1.55 and 0.4 W/K
sweep: sensor accuracy σ_T	0.05, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1 K
near runaway: true α; prior mean and standard deviation	1.92; 1.82 and 0.3 W/K
near runaway: sensor accuracy σ_T	0.4 K
closure-row tolerance σ_{phys}	0.001 K
fit stopping tolerance; iteration limit	10^{-8}; 400
self-check: probe steps; draws; gate; threshold	1, 2; 64; 0.25; 0.5
particle filter: particles; resampling threshold	8,000; 0.5
random seeds: self-check; particle filter	0; 1

the moderate point and 156.25 K per W/K near runaway, 9.77 times larger. The unknowns are $z = (\alpha, T)$, and the rows of Figure 2, each written as a residual divided by its standard deviation, are

$$r_c = \frac{T - T(\alpha)}{\sigma_{\text{phys}}}, \tag{4}$$

$$r_y = \frac{T - y}{\sigma_T}, \tag{5}$$

$$r_p = \frac{\alpha - \mu_0}{s_0}, \tag{6}$$

with the closure tolerance $\sigma_{\text{phys}} = 0.001$ K, the reading y and its accuracy σ_T, and the prior mean μ_0 and standard deviation s_0 of Table 1. Table 5 in the appendix collects the symbols.

3.2 The probabilistic model

The three rows are factors of the general kind this subsection describes, each with weight one once divided by its standard deviation. 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,

$$p(z \mid y) \propto \prod_k \exp\big(-\tfrac{1}{2}\, w_k\, r_k(z)^2\big). \tag{7}$$

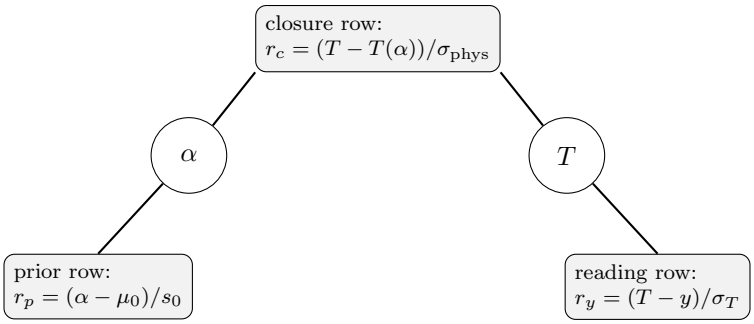


Figure 2: The model behind the estimate. The two circles are the unknowns, the leakage coefficient and the room temperature; the shaded boxes are the rows that tie them, each a residual already divided by its standard deviation. The closure row states the steady heat balance of Eq. (2) to a small tolerance, the reading row ties the room temperature to the sensor, and the prior row states what is known about the leakage coefficient in advance. The leakage coefficient is never read directly: it is known only through the room temperature, and how well depends on the slope of Eq. (2) at its value.

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:

$$C(z) = \tfrac{1}{2} \sum_k w_k \, r_k(z)^2, \tag{8}$$

$$z^* = \arg\min_z C(z), \tag{9}$$

$$z \leftarrow z - (J^\top W J)^{-1} J^\top W \, r(z) \quad \text{until } \|J^\top W \, r(z)\| < \epsilon. \tag{10}$$

Near the minimum the cost (8) 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:

$$p(z \mid y) \approx \mathcal{N}\big(z^*, \Lambda^{-1}\big), \tag{11}$$

$$\Lambda = J^\top W J, \tag{12}$$

$$\sigma_i = \sqrt{(\Lambda^{-1})_{ii}}. \tag{13}$$

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,

$$\log Z = -C(z^*) + \tfrac{d}{2} \log 2\pi - \tfrac{1}{2} \log\det \Lambda + \sum_k \tfrac{1}{2} \log \frac{w_k}{2\pi}. \tag{14}$$

Differences of $\log Z$ between hypotheses, from (14), 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.

3.3 Exact posterior of the leakage coefficient

With one parameter the posterior need not be approximated. The room temperature enters only the closure and reading rows, both quadratic in T; integrating it out, or equivalently minimizing over it, leaves a reading variance that adds the two tolerances, and the log posterior density of α is, up to a constant,

$$\log p(\alpha \mid y) = -\frac{(\alpha - \mu_0)^2}{2 s_0^2} - \frac{\big(T(\alpha) - y\big)^2}{2\,(\sigma_T^2 + \sigma_{\text{phys}}^2)}. \tag{15}$$

The fitted $\hat{\alpha}$ of (9) is the mode of (15), and the Laplace width $\hat{\sigma}_\alpha$ of (13) is the inverse square root of the curvature of $-\log p$ at the mode; to first order in the slope,

$$\hat{\sigma}_\alpha \approx \frac{\sqrt{\sigma_T^2 + \sigma_{\text{phys}}^2}}{dT/d\alpha}. \tag{16}$$

The Laplace band is exact when $T(\alpha)$ is linear across it. The response (2) bends, and its relative bend across one half-width of the band is twice

$$\kappa = \frac{\hat{\sigma}_\alpha}{\alpha^* - \hat{\alpha}}, \tag{17}$$

the band's half-width measured against its distance to runaway; κ is the quantity by which the results are organized.

3.4 Self-check of the Laplace band

The engine's self-check compares the Gaussian band with the exact shape of the fitting cost along α without integrating it. It fixes α at $\hat{\alpha} + m\,\hat{\sigma}_\alpha$ for $m = \pm 1, \pm 2$, refits the room temperature, and compares the rise of the profiled cost C_{prof} with the rise $m^2/2$ a Gaussian predicts; the curvature score is the worst relative mismatch,

$$D = \max_m \frac{\big|C_{\text{prof}}(\hat{\alpha} + m\,\hat{\sigma}_\alpha) - C_{\text{prof}}(\hat{\alpha}) - m^2/2\big|}{m^2/2}. \tag{18}$$

It then draws 64 values of α from the band, weights each by the ratio of the profiled density to the Gaussian density, and reports the effective-sample ratio of Kong et al. (1994),

$$E = \frac{1}{n \sum_i \bar{w}_i^2}, \tag{19}$$

with $n = 64$ draws and normalized weights $\bar{w}_i$. A band passes when $D \le 0.25$ and $E \ge 0.5$ and is flagged otherwise. The check costs 68 profile fits per point; its draws use seed 0.

3.5 Particle filter

The engine's particle filter draws 8,000 values of α from the prior. It does not have to guess at the room temperature: the closure row (4) determines T once α is fixed, so for each draw the filter solves that row for T and then samples T about the solution with the width the row's tolerance allows. Integrating a variable out this way, instead of holding it at a starting value, is what lets

the reading reach the weights at all. Writing $q(T \mid \alpha)$ for the density T is drawn from, the weight of particle i is the density (7) divided by the density it was drawn from,

$$\log w_i = -\tfrac{1}{2}\Big(\frac{T_i - T(\alpha_i)}{\sigma_{\text{phys}}}\Big)^2 - \tfrac{1}{2}\Big(\frac{T_i - y}{\sigma_T}\Big)^2 - \log q(T_i \mid \alpha_i). \tag{20}$$

The prior is absent from (20) because the draw already carries it, and scoring it again would count it twice. Both remaining terms vary from particle to particle, so the reading moves the cloud. Because T enters (4) linearly, the solution is exact in one step and the drawn T_i cancels from (20) algebraically, which leaves the marginal density of α alone: the filter estimates the same one-parameter posterior (15) the quadrature integrates, by an independent route. When the effective sample size of the weights falls below 0.5 of the particle count the cloud is resampled systematically. Distinct particles are counted after rounding α to 6 decimals; the filter uses seed 1, and each particle's solve stops at a step below 10^{-12} or 400 iterations.

3.6 Numerical integration

The density (15) is integrated by the trapezoid rule. A grid of 20,000 intervals over the prior mean ± 12 prior standard deviations below α^* locates the window within 60 nats of the peak, and 200,001 points span that window; the density is evaluated in closed form, the grid supplies the normalizing mass and the moments, each band or tail probability is integrated on its own grid over the shorter side of the window, and quantiles are refined by Newton steps. The integration is accepted only with at least 500 grid steps per posterior standard deviation, window edges at least 50 nats below the peak, and posterior probability beyond α^* below 10^{-7}; every point met all three, and a grid twice as fine leaves every printed value unchanged.

3.7 Scaling of the heat load

By (2) every temperature difference is proportional to Q_0, so (15) depends on α only through

$$\frac{T(\alpha) - y}{\sigma_T} = \frac{Q_0}{\sigma_T}\Big(\frac{1}{G - \alpha} - \frac{1}{G - \alpha_{\text{true}}}\Big). \tag{21}$$

Multiplying Q_0, σ_T and σ_{phys} by the same factor leaves every statement about α unchanged. The whole study is run a second time with 10 times the heat and the same factor on every accuracy and tolerance, as a check of the implementation against this property.

3.8 Protocol

At the moderate point the fit, its band, the self-check and the particle filter are run at each of the eight sensor accuracies; at the point near runaway they are run once. The fit starts from the prior mean and stops at a tolerance of 10^{-8} or 400 iterations. Every band is then graded against the integral of (15), and the study is repeated at the scaled heat load of Section 3.7.

3.9 Metrics

Whether the self-check draws a boundary is read from the curvature score D of (18) against its gate 0.25 and from the effective-sample ratio E of (19) against 0.5; a boundary is clean when the verdict changes once along the sweep. Whether the band is right is read from the exact posterior. The width ratio $\hat{\sigma}_\alpha/\text{sd}$ compares the Laplace width with the exact standard deviation and is one

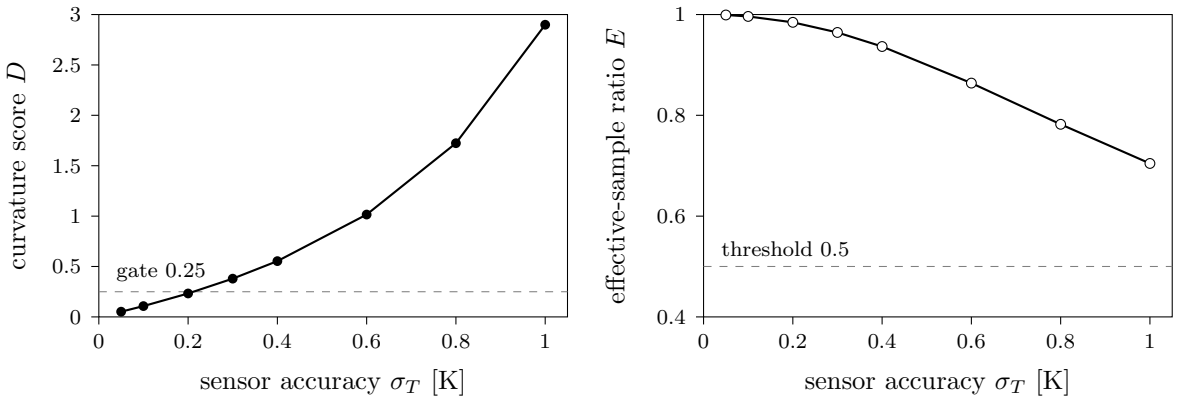


Figure 3: The self-check across the sensor-accuracy sweep at $\alpha = 1.75$ W/K. Left: the curvature score of Eq. (18) at each of the eight accuracies, with its gate dashed; a band is flagged when the score exceeds the gate. Right: the effective-sample ratio of Eq. (19), with its threshold dashed. The score rises monotonically and crosses the gate once, between 0.2 K and 0.3 K; the ratio stays above its threshold everywhere, so the score alone decides every verdict.

for a correct band. The band probabilities P_{68} and P_{95} are the exact probabilities inside $\hat{\alpha} \pm \hat{\sigma}_\alpha$ and $\hat{\alpha} \pm 1.96\,\hat{\sigma}_\alpha$, against their nominal 68.3 % and 95 %; they are the measure that matters to a reader who acts on the band. The exact probability below and above the 95 % band, against 2.5 % each, says on which side the band errs, which near a runaway condition decides whether the error is conservative. The skewness of the exact posterior, zero for a Gaussian, and the offset of $\hat{\alpha}$ from the exact mean in exact standard deviations describe the shape the band misses. The relative band width κ of (17) is the explanatory variable. For the particle filter the mean and standard deviation of the cloud are compared with the exact mean and standard deviation, in units of the exact standard deviation, which is the comparison that says whether it can serve as a second reference; the number of distinct particles and the effective sample size before resampling report what the sharp likelihood costs the cloud.

4 Results

4.1 Self-check across the sensor accuracy

Figure 3 plots the two scores of the self-check against the sensor accuracy at the moderate operating point, the curvature score on the left and the effective-sample ratio on the right, each with its threshold dashed. The curvature score rises monotonically, from 0.0518 at 0.05 K to 2.898 at 1 K, and crosses the gate once: it is 0.233 at 0.2 K and 0.379 at 0.3 K. The effective-sample ratio falls from 0.999 to 0.705 and never reaches its threshold of 0.5, so the curvature score decides every verdict: three bands pass and five are flagged.

4.2 Exact and approximate posteriors

Table 2 sets the fit and its band beside the exact posterior at all nine points, and Figure 4 draws the two densities at four of them on the band's standardized axis. At 0.05 K the Laplace width is 0.999 of the exact standard deviation and the exact skewness is -0.075: the two curves in the top left panel are indistinguishable. At the first flagged accuracy, 0.3 K, the width ratio is 0.947

Table 2: The Laplace band at every study point against the exact posterior of Eq. (15). Rows one to eight are the sweep at $\alpha = 1.75$ W/K; the starred row is the operating point near runaway at $\alpha = 1.92$ W/K. $\hat{\alpha}$ and $\hat{\sigma}_\alpha$ are the fit and its Laplace width, sd the exact standard deviation, D the curvature score and P_{95} the exact probability inside the nominal 95 % Laplace band. Widths and means in W/K.

σ_T [K]	$\hat{\alpha}$	$\hat{\sigma}_\alpha$	exact mean	exact sd	$\hat{\sigma}_\alpha$/sd	skewness	D	P_{95}	verdict
0.05	1.7500	0.0031	1.7499	0.0031	0.999	-0.08	0.052	0.949	pass
0.1	1.7500	0.0063	1.7495	0.0063	0.994	-0.15	0.108	0.948	pass
0.2	1.7498	0.0125	1.7479	0.0128	0.977	-0.31	0.233	0.942	pass
0.3	1.7496	0.0188	1.7451	0.0199	0.947	-0.48	0.379	0.931	flag
0.4	1.7492	0.0251	1.7410	0.0278	0.903	-0.68	0.553	0.916	flag
0.6	1.7482	0.0379	1.7275	0.0491	0.772	-1.30	1.016	0.873	flag
0.8	1.7468	0.0509	1.7026	0.0854	0.596	-2.10	1.724	0.813	flag
1	1.7449	0.0642	1.6640	0.1352	0.475	-2.13	2.898	0.744	flag
0.4*	1.9200	0.0026	1.9198	0.0026	0.991	-0.19	0.141	0.947	pass

and the skewness -0.48. At 1 K the band is 0.475 of the exact width and the skewness is -2.13; the exact density in the bottom left panel carries a long tail toward small α, its 2.5 % quantile at 1.298 W/K against the band's 1.619 W/K. The band's center also sits above the exact mean, by 0.037 exact standard deviations at the most accurate sensor and 0.599 at the least accurate: the fit is the mode, and the mode of a density skewed toward small values lies above its mean.

4.3 Probability inside the band

Figure 5 reads the band the way a user reads it, as a statement of probability. On the left, the exact probability inside the 95 % and the 68.3 % bands is plotted against the sensor accuracy with the nominal values dashed; on the right, the exact probability that falls outside the 95 % band is split into the side below it and the side above it. Wherever the self-check passes, P_{95} is between 0.9417 and 0.9495, within 0.8 percentage points of nominal; wherever it flags, P_{95} is between 0.7443 and 0.9312, 1.9 to 20.6 points short. The gate therefore separates bands that are nearly right from bands that are measurably wrong, and the verdict agrees with the exact reference at all nine points. The shortfall is one-sided. The probability above the band, toward runaway, is at most 2.1 % at any point, below the nominal 2.5 %, and falls to 1.4×10^{-5} at 1 K; the probability below the band rises from 3.0 % to 25.6 %. The Laplace band overstates how close to runaway the coefficient may be and understates how far below the fit it may lie.

4.4 Operating point near runaway

At $\alpha = 1.92$ W/K, with the 0.4 K accuracy that the moderate point flags, the fit is $\hat{\alpha} = 1.91999$ with a band of 0.00256 W/K, the curvature score is 0.141 and the band passes (bottom right of Figure 4 and the starred row of Table 2). The exact reference agrees: the width ratio is 0.991 and $P_{95} = 0.947$. The steep response near runaway narrows the band faster than the approach to runaway shortens the distance it must stay clear of: κ of (17) is 0.0320 here, against 0.0500 at the last passing and 0.0750 at the first flagged sweep point. Ordered by κ, the curvature scores of all nine points, sweep and near runaway together, are in increasing order, so the band's width relative to its distance to runaway, not the proximity to runaway itself, sets its reliability.

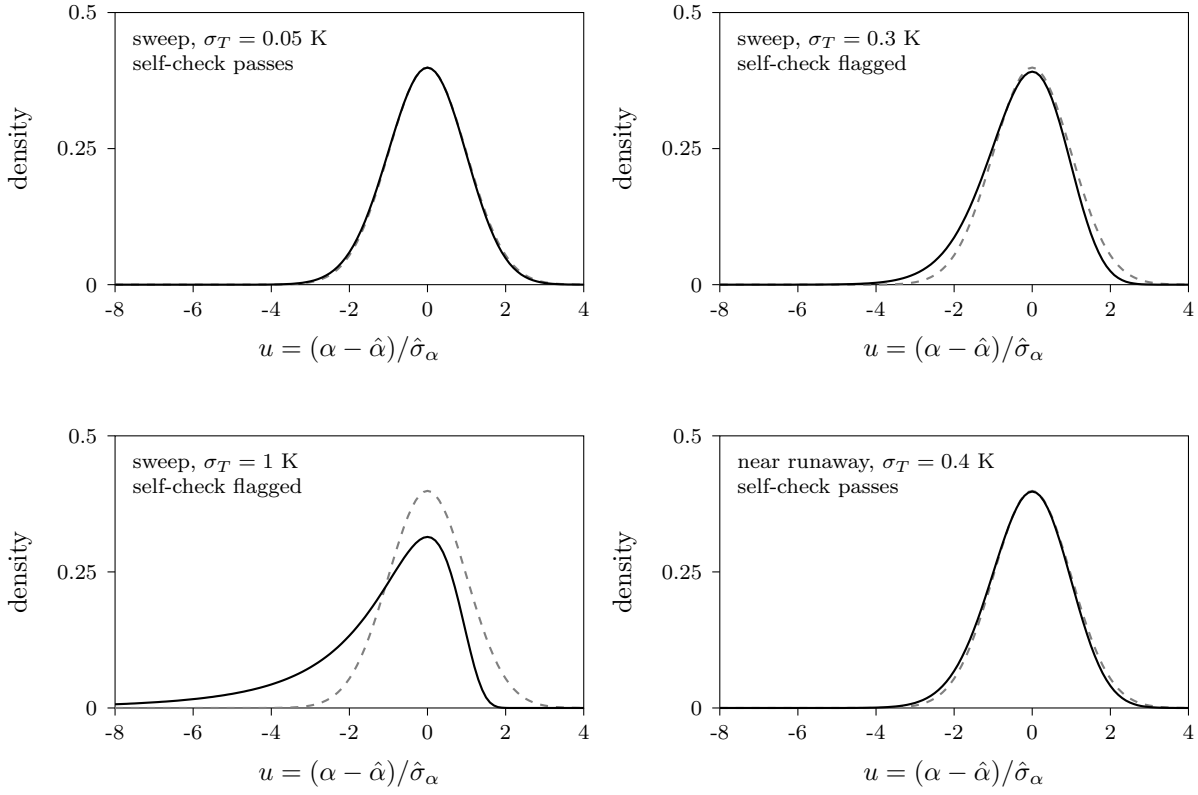


Figure 4: The exact posterior density of the leakage coefficient (solid) and its Laplace approximation (dashed), on the standardized axis u of the Laplace band, on which the approximation is the standard normal density; the exact density is multiplied by $\hat{\sigma}_\alpha$ so that both integrate to one on this axis. Top left: the most accurate sensor of the sweep. Top right: the first accuracy the self-check flags. Bottom left: the least accurate sensor. Bottom right: the operating point near runaway. The exact density leans toward small α — away from runaway — and the lean grows with the sensor noise, while its right flank stays close to the normal curve.

4.5 Particle filter

Table 3 lists the particle filter's result at every point and Figure 6 sets its width beside the exact and the Laplace widths. The filter agrees with the integral everywhere: its mean is within 0.19 of an exact standard deviation of the exact mean at every one of the 9 points, and its width is within 6 % of the exact width. At the most accurate sensor it reports 1.74967 W/K against an exact 1.74987, and at 1 K it reports 1.66320 against an exact 1.66395. Near runaway it reports 1.91926 against an exact 1.91975.

The two estimates are independent in construction: the quadrature integrates the one-parameter density on a grid, while the filter samples the two-variable density and lets the weights do the marginalizing. Agreement between them is therefore evidence about the answer rather than about either method, and it is what makes the band's failure legible. At 1 K the Laplace width is 2.1 times narrower than both, so the disagreement in Figure 6 at the noisy end is the band's, not the reference's.

What the filter costs is diversity. Its particles come from the prior, which at the most accurate sensor is far wider than the posterior, so only 68 of 8,000 particles carry appreciable weight there

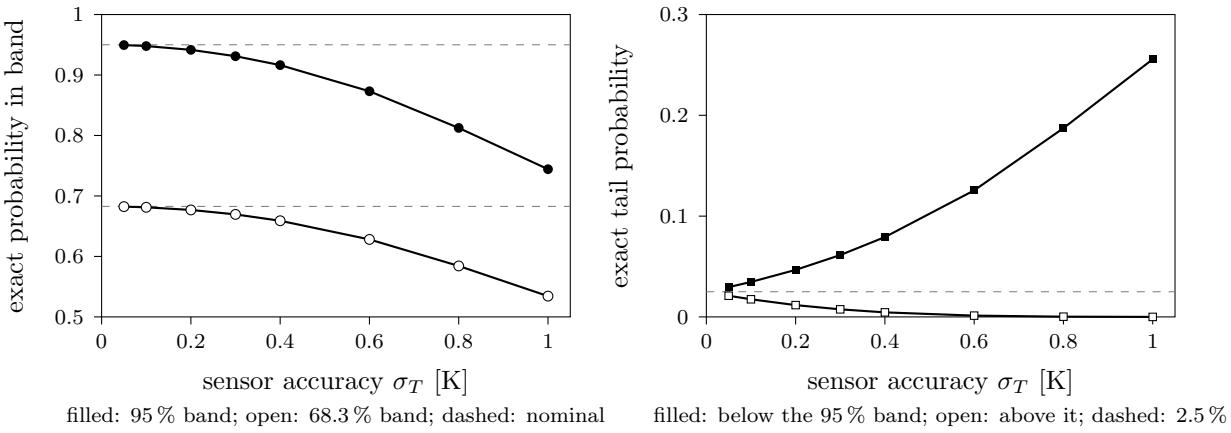


Figure 5: How much exact probability the Laplace bands hold across the sweep. Left: the exact probability inside the $\pm 1.96\,\hat{\sigma}_\alpha$ band (filled) and the $\pm\hat{\sigma}_\alpha$ band (open), against their nominal values (dashed). Right: the exact probability that falls below the 95 % band, toward small α (filled), and above it, toward runaway (open), against the nominal 2.5 % per side (dashed). The shortfall of the band is almost entirely on the side away from runaway.

and 1.8 % of the cloud survives resampling as distinct values; by 1 K, where the posterior is wide, that rises to 2326 particles and 40.9 %. A sharp posterior is expensive to sample from the prior; it is not a statement that the estimate is wrong.

4.6 Invariance under scaling of the heat load

Table 4 compares the study with the run at 10 times the heat. The first flagged accuracy scales by the same factor, from 0.3 K to 3 K, the verdicts at all nine points and every particle count are equal, and all 63 continuous quantities agree to a relative 10^{-6}, as (21) requires. Every room temperature any fit, profile fit or particle visits lies between 300.9 and 313.8 K in this report's normalization.

5 Discussion

The self-check is right at every point this study can grade: its gate falls between bands that hold within 0.8 points of their nominal 95 % and bands that fall short by 1.9 points or more. The grading has limits. The model has one parameter and one reading, where the profiled cost the self-check probes is the exact marginal; with several correlated parameters the probe follows one direction at a time and can miss a bend across directions. The reading is noise-free, so the nine points are nine evaluations of one likelihood shape each, not a coverage study over noise draws; a noisy reading moves the fit along the curve and with it κ. The gate 0.25 and the probe steps are the engine's settings, checked here and not calibrated; a different model could put the same score at a different band probability. The prior is broad and Gaussian, and a prior that ends at α^* would change the shape of the posterior on the runaway side, where this study finds the band conservative.

The direction of the error matters for use. Because the response flattens away from runaway, a reading pins large α better than small, and the exact posterior leans toward small values. A Laplace band on this model therefore errs toward caution about runaway and toward overconfidence that α is not small. That is a property of this response, not of the approximation in general; a response that steepens away from the quantity of concern would reverse it.

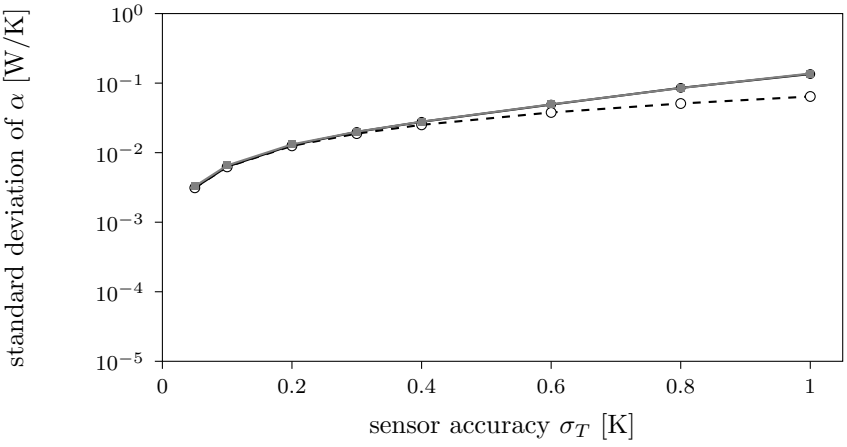


Figure 6: The width of the estimate of α reported by each method across the sweep, on a logarithmic axis. The exact posterior standard deviation and the Laplace width agree at the accurate end and part as the sensor noise grows. The particle filter's width tracks the exact one at every accuracy, by a route that shares nothing with the quadrature, so the gap that opens at the noisy end belongs to the band alone (Section 4.5).

The filter's agreement with the integral is not free, and its cost is the one Snyder et al. (2008) describe. Its particles come from the prior, so where the posterior is much narrower than the prior – the accurate end of the sweep – only 68 of 8,000 particles carry appreciable weight, and the estimate rests on them. That is a statement about the price of sampling a sharp likelihood from a broad proposal, not about the answer; it is why the effective sample size is reported beside every mean in Table 3. A proposal centered on the Laplace band would be cheaper, and on this model it would also be wrong: the band is the object under test, and where it is too narrow a sampler drawn from it inherits that error instead of exposing it. On a model with several parameters the prior becomes an expensive proposal quickly, and the filter would need a better one to stay practical.

6 Conclusion

This report asked when a Gaussian uncertainty band on a thermal-runaway parameter misstates the uncertainty, and whether a curvature self-check detects it, for a room whose leakage coefficient is estimated from one temperature reading. Graded against the exact posterior, the band holds its nominal 95 % to within 0.8 points while its half-width is small compared with its distance to runaway and falls to 0.744 when it is not. The self-check passes and flags exactly those bands: at the moderate operating point its boundary lies between 0.2 K and 0.3 K of sensor accuracy. Near runaway the band is narrow and correct ($P_{95} = 0.947$), because the steep response pins the coefficient. Where the band errs, it errs away from runaway: at most 2.1 % of the probability lies above it, up to 25.6 % below. The engine's particle filter, which integrates the room temperature out of each particle, reproduces the exact posterior at every point by a route that shares no machinery with the quadrature, so a flagged band has a second estimate to fall back on and the flag can be read as an instruction to switch methods rather than only as a warning. What remains is a grading of the self-check on models with several parameters and noisy readings, where the profiled cost it probes is no longer the exact marginal and the prior is no longer a cheap proposal.

Table 3: The particle filter at every study point, with 8,000 particles, beside the exact posterior it is compared with. The two are computed by different routes – the filter samples the prior and weights, the exact values integrate the density of Eq. (15) on a grid – and agree at every point. ESS is the effective sample size of the weights before resampling: it measures how much of the cloud the sharp likelihood leaves alive, and is smallest where the posterior is narrowest. Means and widths in W/K; the starred row is the operating point near runaway.

σ_T [K]	true α	filter mean	exact mean	filter sd	exact sd	ESS
0.05	1.75	1.74967	1.74987	3.28×10^{-3}	3.13×10^{-3}	68
0.1	1.75	1.74953	1.74948	6.55×10^{-3}	6.29×10^{-3}	146
0.2	1.75	1.74796	1.74789	1.31×10^{-2}	1.28×10^{-2}	307
0.3	1.75	1.74557	1.74514	2.00×10^{-2}	1.99×10^{-2}	470
0.4	1.75	1.74172	1.74104	2.79×10^{-2}	2.78×10^{-2}	642
0.6	1.75	1.72807	1.72746	4.95×10^{-2}	4.91×10^{-2}	1052
0.8	1.75	1.70266	1.70257	8.51×10^{-2}	8.54×10^{-2}	1598
1	1.75	1.66320	1.66395	1.38×10^{-1}	1.35×10^{-1}	2326
0.4*	1.92	1.91926	1.91975	2.75×10^{-3}	2.58×10^{-3}	72

Table 4: The study at the equipment heat of this report and at ten times that heat, with every temperature accuracy and tolerance scaled by the same factor. All 63 continuous quantities compared (fit, width, curvature score, effective-sample ratio, filter width and share of distinct particles at each of the nine points) agree to a relative 10^{-6} with an absolute floor of 10^{-9}; verdicts and particle counts are equal.

quantity	this report	ten times the heat
equipment heat Q_0 [W]	1	10
sensor accuracies σ_T [K]	0.05–1	0.5–10
closure-row tolerance [K]	0.001	0.01
first flagged accuracy [K]	0.3	3
curvature score, most and least accurate	0.0518, 2.8984	0.0518, 2.8984
Laplace width [W/K], most and least accurate	0.00313, 0.06421	0.00313, 0.06421
near runaway: Laplace width [W/K]; score	0.00256; 0.1412	0.00256; 0.1412
distinct particles, sweep; near runaway	140; 172	140; 172

A Reproduction

From the repository root, after changing to the study directory, the first command computes every number of this report into the study's results file and the second writes this document and builds its PDF:

```
cd docs/terra_graph_engine/case_studies/level2_room_hall/when_laplace_lies
PYTHONPATH=../../../../../src python run.py
PYTHONPATH=../../../../../src python make_report.py --pdf
```

The first command accepts `--out` for another directory and `--fast`, which reads the scaled-heat run from the committed results file instead of repeating it. The computation is deterministic: the self-check uses seed 0 and the filter seed 1, and the same command writes the same file. The acceptance test is `tests/terra_graph_engine/prototype/test_when_laplace_lies_study.py`; it regenerates the results, compares them with the committed file, and rebuilds this document byte for byte.

B Data and provenance

All data are synthetic and generated by the study; no external data set is used. The model, the fit, the self-check and the particle filter are those of the engine at the commit this report was built from; the exact posterior and every grade are computed by the study itself. The equipment heat is $Q_0 = 1$ W in Eq. (2). The model at commit `5be377ed` used $Q_0 = 10$ W, with sensor accuracies and the closure tolerance ten times larger; its figures in kelvin are not comparable. An earlier generation of this report also carried different filter numbers: the engine's filter then held the variables without a prior at their starting values, and reported a cloud on the prior mean (1.54993 W/K, standard deviation 1.83×10^{-4} W/K) at every sensor accuracy. Those values are recorded in the study's results file under `supersedes_filter`; every fit, self-check and exact-posterior number in this report is unchanged from that generation.

C Notation

Table 5 lists the symbols.

Table 5: The symbols of this report.

symbol	meaning	unit
α	leakage coefficient, the unknown parameter	W/K
α^*	runaway value of α, equal to G	W/K
T, T_s	room temperature; cooling-supply temperature	K
G	conductance from the room to the supply	W/K
Q_0	equipment heat independent of temperature	W
y, σ_T	the reading; its stated accuracy	K
σ_{phys}	tolerance of the closure row	K
μ_0, s_0	prior mean and standard deviation of α	W/K
$\hat{\alpha}$, $\hat{\sigma}_\alpha$	fitted α and its Laplace width	W/K
u	$(\alpha - \hat{\alpha})/\hat{\sigma}_\alpha$, the standardized axis	—
D	curvature score of the self-check	—
E	effective-sample ratio of the self-check	—
κ	$\hat{\sigma}_\alpha/(\alpha^* - \hat{\alpha})$, band width relative to distance to runaway	—
P_{68}, P_{95}	exact probability inside the 68.3 % and 95 % Laplace bands	—

References

[Bates and Watts 1980] Bates, D. M., Watts, D. G. (1980). Relative curvature measures of nonlinearity. Journal of the Royal Statistical Society, Series B 42(1), 1–16. doi:10.1111/j.2517-6161.1980.tb01094.x.

[Gordon et al. 1993] Gordon, N. J., Salmond, D. J., Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings F (Radar and Signal Processing) 140(2), 107–113. doi:10.1049/ip-f-2.1993.0015.

[Kong et al. 1994] Kong, A., Liu, J. S., Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. Journal of the American Statistical Association 89(425), 278–288. doi:10.1080/01621459.1994.10476469.

[Rue et al. 2009] Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society, Series B 71(2), 319–392. doi:10.1111/j.1467-9868.2008.00700.x.

[Snyder et al. 2008] Snyder, C., Bengtsson, T., Bickel, P., Anderson, J. (2008). Obstacles to high-dimensional particle filtering. Monthly Weather Review 136(12), 4629–4640. doi:10.1175/2008MWR2529.1.

[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.

[Vassighi and Sachdev 2006] Vassighi, A., Sachdev, M. (2006). Thermal runaway in integrated circuits. IEEE Transactions on Device and Materials Reliability 6(2), 300–305. doi:10.1109/TDMR.2006.876577.

[Vehtari et al. 2024] Vehtari, A., Simpson, D., Gelman, A., Yao, Y., Gabry, J. (2024). Pareto smoothed importance sampling. Journal of Machine Learning Research 25(72), 1–58. url:https://jmlr.org/papers/v25/19-556.html.

[Yao et al. 2018] Yao, Y., Vehtari, A., Simpson, D., Gelman, A. (2018). Yes, but did it work?: Evaluating variational inference. Proceedings of the 35th International Conference on Machine Learning, PMLR 80, 5581–5590. url:https://proceedings.mlr.press/v80/yao18a.html.

[Zapater et al. 2015] Zapater, M., Tuncer, O., Ayala, J. L., Moya, J. M., Vaidyanathan, K., Gross, K., Coskun, A. K. (2015). Leakage-aware cooling management for improving server energy efficiency. IEEE Transactions on Parallel and Distributed Systems 26(10), 2764–2777. doi:10.1109/TPDS.2014.2361519.