The Cost of a Bit in a World: Landauer Meets the Family Law at a Pure Number

Martin Scholl — Independent Researcher

It Is All One — Information Science Notes. July 2026 (working draft)

Abstract. This note reports one exact identity, one quarantined near-miss, and a set of leads for somebody who works on information systems professionally. The identity: combine Landauer's 1961 result that erasing one bit costs $k_B \cdot T \cdot \ln 2$ of heat with this series' family law, $k_B \cdot T = \hbar c/2\pi R$, and the temperature cancels. What remains is that erasing one bit inside a world of curvature radius R costs exactly $\ln 2/2\pi = 0.1103$ of that world's own entry-door energy $\hbar c/R$ — a pure number, dimensionless, temperature-free, and the same on every rung of the ladder. Landauer and P2 meet at a constant. Read the other way, this makes the chain the author has been pressing — information is energy, energy is mass, mass is condensed space — tight rather than rhetorical: its one loose link is Landauer's temperature dependence, and the family law removes it the moment one stands on a rung. The near-miss: the colour rung's bit-cost is 107.2 MeV against the muon's 105.658 MeV, 1.49% apart, which would bear on the lepton tower named as this corpus's fourth open item. It is quarantined here for two stated reasons — one landing in four rungs at that precision is what chance delivers, and there is no reason a bit-erasure cost on the colour rung should be a lepton mass. The remainder of the note surveys where the framework and the science of information systems already touch: the Fisher information metric is a curvature theory of learning that says, independently, what P2 says; and the gearbox conjecture — the α calculation this framework is waiting on — is a bit-counting problem in its own statement. Every citation is from memory and none has been verified; the whole note fails caveat (ix) by construction. It is offered to information scientists exactly as the chemistry notes are offered to chemists: check the numbers on a calculator, then decide whether the convergence is coincidence or geometry.

1. One Number a Reader Can Check

Three ingredients, none of them new. From 1961: erasing one bit of information costs at least $k_B \cdot T \cdot \ln 2$ of heat, where k_B is Boltzmann's constant and T the temperature of the surroundings — Landauer's principle, since measured on single particles. From this series' Section 4: a world of curvature radius R has temperature $k_B \cdot T = \hbar c/2\pi R$. From P2: that world's entry door costs $E_{door} \approx \hbar c/R$.

Substitute the second into the first and the temperature disappears:

$$E_{bit} = k_B \cdot T \cdot \ln 2 = (\ln 2 / 2\pi) \cdot \hbar c/R = 0.110318 \cdot E_{door}$$

Erasing one bit inside a world costs eleven percent of that world's door, exactly, on every rung. No temperature, no material, no scale. A bit is a fixed fraction of the price of admission.

That is the whole result, and it takes a calculator: ln2 = 0.693147, divided by 2π = 6.283185, gives 0.110318. The rest of this note is context and leads.

2. The Chain, and Its One Loose Link

The author's proposition runs: information is energy, energy is mass, mass is condensed space. Stated with its coefficients, the chain is uneven.

Energy to mass is exact and coefficient-free, $E = mc^2$. Mass to condensed space is P2 — this series' postulate, not established physics, adopted with Clifford's thesis and Weyl's repair. Information to energy is the interesting one: it is not an identity but an interconversion, at the rate $k_B \cdot T \cdot \ln 2$ per bit, and it carries a temperature. That is the loose link. $E = mc^2$ is a fixed exchange rate; Landauer's is not.

Section 1 is the observation that this series already owns the thing that tightens it. The family law trades temperature for curvature radius, so once an observer is standing on a rung the exchange rate is fixed by geometry alone. One bit becomes ln2/2π of a door — and by P2 a door is a winding radius, which is to say a quantity of condensed space. The chain closes, and closes on a pure number.

3. Rung by Rung — and One Landing That Must Be Quarantined

Applying the identity across the ladder, using each rung's own curvature radius: chemistry (1 μm) has a door of 197.3 meV and a bit-cost of 21.8 meV; the electron's Compton cell (3.86×10^{-13} m) a door of 511 keV and a bit-cost of 56.4 keV; the colour world (0.203 fm) a door of 972.1 MeV and a bit-cost of 107.2 MeV; the weak bubble (2.45×10^{-18} m) a door of 80.5 GeV and a bit-cost of 8.89 GeV.

One of those lands near something. The colour rung's 107.2 MeV sits 1.49% from the muon's 105.658 MeV — and the muon is precisely the candidate ground tone of the lepton tower, this corpus's fourth open item and a named falsifier.

It is quarantined, and the reasons are worth stating rather than gesturing at. First, arithmetic: one landing in four rungs at one-and-a-half percent against a dense particle spectrum is roughly what chance produces, and the other three rungs land on nothing within a few percent. Second, and worse: the coincidence is unmotivated. There is no reason for a bit-erasure cost computed on the colour rung to be a lepton mass. Until something explains why those two quantities should be the same object, this belongs on the card beside 1/2e and the geometric mean — and the geometric mean, tested yesterday, failed.

4. What the Framework Finds When It Looks at a Neural Network

The organising question behind these notes was whether the framework's claim — one geometric object at many magnifications — reaches an information system. Taking the framework's own four terms in turn, and reporting honestly:

Algebra: yes, and non-commutativity is load-bearing. The core operation is matrix multiplication, associative but order-sensitive, and attention makes the asymmetry explicit — a head computes $Q \cdot K^T$, so position i attending to j is not j attending to i. Order matters for the reason it matters in the quaternions: the operation is a directed relation. But the algebra is associative and has no division; it is nowhere on the Cayley–Dickson ladder, and Cayley's theorem has no purchase on it.

Differential equations: genuinely, not metaphorically. The residual stream updates as $x(\ell+1) = x(\ell) + F(x(\ell))$, which is Euler integration of $dx/d\ell = F(x,\ell)$ at unit step. Depth is an integration variable and the residual stream is a field on layer-time and token-position. (Lead: the Neural ODE literature.)

Spaces: yes, and one echoes P2. Features appear to be directions in a shared vector space of order 10^3–10^4 dimensions, and there are far more features than dimensions, packed at near-orthogonal rather than orthogonal angles — 'superposition'. The logic is P2's affordability argument: bending the open directions is expensive, so the structure is carried by packing. More things than room, made affordable by geometry.

Things interacting: attention heads, but the interaction is non-local. Any position reaches any other in one step — no delay, no neighbourhood, and no seal. That is the sharpest structural difference from a framework in which exit cones and horizons are foundational.

5. What It Does Not Find — Three Absences, One Fatal

No conservation law, and this is the fatal one. This framework's predictive power comes almost entirely from books that must balance: the Noether ledger, redshift as transfer and never write-off, the winding number that cannot unwind, eternity as topological protection. An information system has no such invariant — layer normalisation destroys scale information at every step, there is no variational principle, no continuous symmetry of a dynamics, therefore no Noether theorem. Things are computed and discarded.

No metric that responds to content. The activation space has an inner product but it does not change because something is present. Nothing bends. (Section 6 qualifies this: there is a responsive metric, but it lives in parameter space.)

Time is not a symmetry — with one exception worth recording. Every layer carries its own weights, so layer forty is not layer four advanced; it is a different apparatus, and the autoregressive mask gives a degenerate light cone, unbounded toward the past and sealed toward the future. But at inference the weights do not move while the residual stream advances: the stage is fixed, the play ticks. That is P3, exactly. It fails during training, when the stage itself moves. Which suggests the useful statement is that learning is precisely the regime in which the stage is not eternal.

6. Two Places the Subjects Already Touch

The first is a curvature theory of learning, roughly forty years old, that nobody appears to have set beside P2. The parameter space of a statistical model carries the Fisher information metric g_ij — equivalently, the second derivative of the Kullback–Leibler divergence between nearby models, which is to say the curvature of the statistical manifold, a genuine Riemannian metric with genuine curvature. And natural gradient descent, the update $g^{-1}\cdot\nabla L$, is the statement that what looks like a force in coordinates is curvature in the metric. Learning is not a push down a landscape; it is flow in a curved space, and the apparent force is an artefact of the wrong coordinates. That is P2, reached independently, in a domain with no gravity in it. (Lead: Amari.)

The second runs the other way, and is the more important of the two. The expectation was that this framework might be implemented in information science. But the engine of Part IV counts entries on a wall — one per Planck area, $\delta Q = T\cdot dS$, Bekenstein's bookkeeping — and that is the step that makes the gravitational coupling come out computed rather than chosen. It is not physics dressed in information language; it is information language, doing the load-bearing work. More pointedly, the gearbox conjecture — named in §11 as the single most valuable calculation this framework points to — is stated in the corpus's own words as: count the circle fiber's seals in their natural pixel and see whether α falls out. That is a bit-counting problem. Not analogous to one. One.

Conjecture, offered as this note's central one: if α ever falls out of counting seals, nothing will have been imported from information science. It will have turned out that the two were one subject at the bottom — which is what this series' title claims and has not yet had to cash.

7. The Compute Conjecture, and Its Honest Ledger

The author's conjecture: the computing centres now being built will prove unnecessary once one can take a second derivative of information flow and find the downhill, the natural curvature, faster than brute force.

What is right about it: the second derivative of information flow is the Fisher metric of Section 6, and using it to find the downhill is the natural gradient. This is already the frontier rather than a speculation — the optimizer that trained most current systems is a diagonal preconditioner, which is a crude curvature approximation, and better ones have been taking ground. (Leads: K-FAC, Shampoo, SOAP, Muon.)

The obstruction is dimensional, not conceptual. Full curvature is an $N \times N$ object; at 10^{12} parameters that is 10^{24} entries, which cannot be formed, stored or inverted at any scale of hardware. Every practical method is therefore a guess at the metric's structure, and realised gains have been factors of two or three because per-step cost eats per-step gain.

Where the conjecture actually bites, and this is its best form: the question is not whether to use the curvature but whether the curvature has structure that makes it cheap — and P2 supplies a candidate property. A world does not have a curvature profile; it has a curvature. If the learning manifold had that property in the right variables, one would get second-order convergence at first-order cost. Nobody knows. It is well posed.

Two counterweights, so the conjecture is not flattered. Training is a shrinking fraction of deployed compute and no optimizer touches a forward pass, so total success cuts one term and not the sum. And the 'bitter lesson' records that for decades every attempt to build in structure has lost to scale plus general methods — imperfectly rebutted by the observation that the lesson concerns hand-crafted structure, and that Newton did not lose to brute-force simulation of planets.

The strongest version, which the author did not quite state: a human acquires language from of order 10^8 words and these systems require 10^{13}. That four-order gap is not an optimizer-step problem, it is a prior problem — the model does not know what is near what before it begins. And what is near what is a metric. So the geometry being sought would not mainly make the descent faster; it would make far less descent necessary. If a factor of ten thousand is hiding anywhere, it is there.

8. Small State Spaces: the Sixty-Four, the Beiwerk, the Encounter

The author's preference for the I Ging's sixty-four as a manageable engine has a historical hook: Leibniz, corresponding with the Jesuit Bouvet around 1703, recognised that the sixty-four hexagrams in the Fu Xi arrangement are the binary numbers 0 to 63 — six lines, broken or unbroken, six bits. One distinction must be kept, though: sixty-four states is a catalogue, while sixty-four weighted variables is a 64-dimensional continuum, incomparably larger. And what makes the I Ging an engine rather than a list is not the sixty-four but the changing lines — a transition rule. States plus a rule. (The engine the Chinese trusted with time was the sexagenary sixty, ten stems meshed against twelve branches.)

The smallness is defensible rather than merely appealing: Rule 110 has two states and eight rules and is Turing complete; Conway's Life has two states. Tiny kernels iterated with interaction produce unbounded complexity. And a modern instance of exactly the proposal exists — vector-quantised models compress to a discrete codebook of a few hundred entries and generate by transitions among codes. A learned I Ging, with the hexagrams found rather than inherited.

The author's Beiwerk claim — reduce to a small core and the remainder is trimming — is measurable and roughly right. The intrinsic dimension of a task, the smallest random subspace within which training still succeeds, comes out in the hundreds to low thousands for large models; low-rank adaptation works at rank eight, sometimes rank one. The structure underneath really is small. Not sixty-four, but on a logarithmic scale sixty-four and a thousand are neighbours and both are nowhere near 10^{12}.

And his sketch of an encounter — two people on a path resolving sex, then friend-or-foe, then uneasy trust, then a signal of harmlessness, then relaxation, modelled as two state vectors sharing a space with a Taktfrequenz — has more empirical backing than a sketch deserves. The sequence has been filmed: cross-cultural work found the eyebrow flash, a raise of order a sixth of a second, apparently universal, and the smile appears to descend from the primate fear grimace. Interactional synchrony is measured — postural sway, breathing, speech rhythm entraining, rising with rapport. And the programme has already succeeded in its simplest case: the social force model treats pedestrians as particles with repulsive terms and a clock, and reproduces lane formation and bottleneck arching without being told to. Humans as vectors in a space with forces is not fantasy. It works, for movement. What nobody has is a state vector carrying the noble, the drunkard, the one who means harm.

Quarantined, at this series' standard: the genetic code also has sixty-four, because $4^3 = 2^6$. The equality is exact and it has attracted a great deal of nonsense.

9. The Two Missing Things

The algebra. Nobody has it. What is observed is that features add — the old word-vector arithmetic, king minus man plus woman — which would make the algebra abelian: no order-dependence whatever. That is almost certainly incomplete; it holds approximately and fails in ways nobody can characterise. P1 is precisely the objection: if the space describes anything that composes, order must matter somewhere. Finding the non-commutative structure underneath the observed addition is the sharpest open problem in these notes, and there are no good candidates.

The weight scale — the author's question, one noble and one peasant, how do they interact. This probably has an answer and it is a power law: word frequencies follow Zipf's inverse rank over many decades, and feature activations look similarly heavy-tailed, a few directions firing constantly and most almost never. The space is steeply aristocratic, and the exponent is measured and unexplained.

The cleanest example of the field's pre-theoretic condition is the scaling laws: loss falls as a power law in parameters over many orders of magnitude, with no derivation whatsoever. In this series' six columns that is a measured anchor awaiting its theorem — the condition physics was in before somebody found the law under the regularity.

And one obstruction recurs, worth naming once. The I Ging has three thousand years of commentary on what a hexagram means when it changes and not one row of what actually happened afterwards. The encounter has an enormous descriptive ethology and almost no record of how particular encounters went. A dynamics cannot be fitted to labels that were never recorded. That may be the honest reason the I Ging stayed an oracle rather than becoming a physics, and it is the first thing to ask a specialist: if a corpus with an outcome column exists anywhere, several questions below become tractable at once.

10. Carnot's Order

The author's remark that the mathematics of these systems is undiscovered but already in use describes the normal order rather than an anomaly, and the precedent is closer to home than it looks. Newcomen's engine ran in 1712; Carnot's analysis came in 1824 and Clausius's formulation in the 1850s. A hundred and forty years of working machines before the law — and Carnot did not derive it from first principles. He derived it from studying the engines. The theory came out of the machine.

That is a methodological hint, and it points at taking these systems apart rather than at theory first. It is also where the two subjects meet a second time: the engine of Part IV is Clausius, the law that came out of the steam engine, applied to a horizon. If information's law has to be extracted from these machines the same way, then the meeting is twice over — once at $\ln 2/2\pi$, and once in the method.

11. Open Questions, Named

For whoever picks this up, in the order they seem tractable. One: the motif count. Domain-independent recurring circuits are real — induction heads, which implement 'if A was followed by B, expect B after A' and care not at all whether A and B are amino acids or ingredients, form abruptly in training and turn on in-context learning; and the same motifs recur across models trained on different data. How many are there? An empirical question with a number at the end. If small, the compute conjecture strengthens sharply; if it runs to millions, it dies. Two: the non-commutative structure of Section 9. Three: whether the Fisher metric has exploitable structure. Four: whether the 10^4 sample-efficiency gap is a metric. Five: whether somebody who counts bits professionally recognises the gearbox. Six: whether anything at all is conserved. Seven: whether an outcome column exists.

Note that questions three and six are the same question. Find the potential and the geometry becomes cheap, because one stops having to measure the landscape everywhere. Find the potential and something is conserved along its level sets — which is the invariant Section 5 says is missing. The author's word for it was intent.

12. The Invitation, and the Flags

The invitation is the same one the chemistry notes make, and in the same spirit. Section 1 is a two-line calculation: $\ln 2$ divided by 2π. Section 3 is four multiplications. Anyone who works on information systems can check both in five minutes and then judge whether a bit being eleven percent of a world's door is a coincidence of unit conventions or a statement about what information is. This note does not know. It reports the number and its one near-miss, with the near-miss quarantined and the reasons for quarantining it written down.

The flags, without softening. Every reference here is a lead rather than a source: the whole note was written from memory without a literature search, so it fails caveat (ix) by construction, and a third of the attributions should be expected to need correcting. Established elsewhere but unverified: Landauer's principle, the Fisher metric and natural gradient, the residual-stream ODE reading, superposition, scaling laws, intrinsic dimension and low-rank adaptation, induction heads and universality, Zipf's law, Rule 110, Leibniz on the hexagrams, the eyebrow flash, interactional synchrony, the social force model. Conjectured here and labelled: that this framework is already an information theory and the gearbox is where that becomes explicit; that the compute conjecture's real target is inductive bias; that questions three and six are one. Quarantined: the muon at 1.49%, and $4^3 = 2^6$. Honest negatives, recorded so they are not lost: no conservation law, no content-responsive metric in activation space, no time symmetry outside inference, a degenerate causal cone, an observed feature algebra that is abelian and therefore incomplete, and the bitter lesson standing as a decades-long argument against the whole programme.

Nothing in this note is claimed as a result of this series. It is parked here, as the chemistry notes are parked, for a reader who knows the territory.

13. The Sentence

Erasing one bit inside any world of this ladder costs eleven percent of that world's door — Landauer's heat and P2's admission price differing by ln2 over 2π and by nothing else — which is the tightest form the author's chain has yet taken, and which suggests that the borrowing between these two subjects runs the way nobody expected: not a field theory implemented in information, but a field theory whose engine already counts bits on a wall, and whose one outstanding calculation is a bit-counting problem wearing a physicist's coat.

References

R. Landauer (1961), irreversibility and heat generation; L. Szilard (1929); J. D. Bekenstein, Phys. Rev. D 7, 2333 (1973); R. Clausius (1865); S. Carnot (1824); G. W. Leibniz, correspondence with J. Bouvet (~1703). S. Amari, information geometry and the natural gradient. Neural ODEs: Chen and co-authors (~2018). Superposition: Elhage and co-authors (~2022). Induction heads and universality: Olsson, Elhage and co-authors. Scaling laws: Kaplan and co-authors (~2020); Hoffmann and co-authors (~2022). Intrinsic dimension: Li and co-authors (~2018); Aghajanyan and co-authors (~2020). Low-rank adaptation: Hu and co-authors. Curvature-aware optimizers: Martens and Grosse (K-FAC), Shampoo, SOAP, Muon. The bitter lesson: Sutton (2019). Cellular automata: Wolfram; Cook on Rule 110. Vector quantisation: van den Oord and co-authors. Zipf's law: Zipf. Ethology of greeting: Eibl-Eibesfeldt. Interactional synchrony: Condon. Social force model: Helbing and Molnár (~1995). The Yijing, Wilhelm and Legge translations. And the papers of this series: the Allgemeine Feldtheorie (P1–P3, §4 the family law, §11 the gearbox, the six columns, caveat (ix)); The Family Law's Cosmic Rung; Tolman from Staticity; The Core of Metric D; Chasing Shadows; The Anchor's First Test. (All citations from memory; the literature-verification pass applies to every one, and with particular force here.)

Acknowledgment: the chain — information is energy, energy is mass, mass is condensed space — and the instinct that the two subjects meet are the author's. Numerical work and drafting by machine (Claude, Anthropic), from one conversation, without a literature search.