

Negative Work as an Information Credit System: Credit-Adjusted Landauer Bounds from Bias, Correlation, and Geometric Structure

Ian Todd

Sydney Medical School

University of Sydney

`itod2305@uni.sydney.edu.au`

Abstract

Landauer’s principle is widely paraphrased as “erasing a bit costs $k_B T \ln 2$ ”. The correct statement is entropic: the minimal work required for logically irreversible state reset scales with the *entropy removed* from the register. This matters because real biological and physical substrates are not featureless thermal baths: they contain *bias* (nonuniform priors), *correlations* (side information), and *geometric structure* (curved or coherent state-space dynamics) that reduce the entropy that must be expelled as heat during code formation.

We propose a bookkeeping view in which such structure constitutes an **information credit** that can be accrued (by creating bias, correlations, or coherent low-entropy order) and later spent to perform information-processing steps with reduced dissipation, including episodes of apparent “negative Landauer work” (net work extraction). We formalize a simple credit ledger that extends standard information-thermodynamic inequalities by adding a redeemable structure term, and we provide toy examples

showing how (i) a tilted register lowers erasure work in proportion to Shannon entropy, and (ii) mutual information with an environment reduces erasure work to a conditional entropy. These examples clarify how “cheap observation” arises when the system already “knows” the answer: measurement cost scales with surprise, not representation size.

Keywords: Landauer principle; nonequilibrium free energy; mutual information; information geometry; code formation; dimensionality

1 Introduction

Episodes of ultra-low dissipation and work extraction in information-processing systems are often described informally as “violations” or “evasions” of Landauer’s limit. The recurrent confusion is semantic: Landauer’s bound is not a fixed price per physical bit, but a lower bound proportional to the entropy reduction of a register [Landauer, 1961, Bennett, 1982, Parrondo et al., 2015].

In practice, code formation and measurement do not occur in a featureless thermal bath. Living systems, engineered controllers, and coherent physical substrates maintain structure across time: *biased priors* (tilted distributions), *correlations* with an environment (side information), and *coherent geometric constraints* that effectively pre-shape the control landscape. Such structure can be treated as a consumable resource. When it is present, the same logical operation can be performed with reduced dissipation; when it is consumed, the reduction must be paid back by increasing entropy elsewhere, consistent with the second law.

This paper offers a compact formalism for that intuition: **negative work as an information credit system**. The intent is not to introduce new thermodynamics, but to provide a transparent ledger linking: (i) entropy removed from a register, (ii) correlations used as side information, and (iii) pre-existing geometric/coherent structure in the control landscape. We focus on minimal models that are easy to simulate and that generate figure-ready predictions.

¹

¹Relevant experimental anchors include biased/tilted erasure protocols [Bérut et al., 2012] and information-

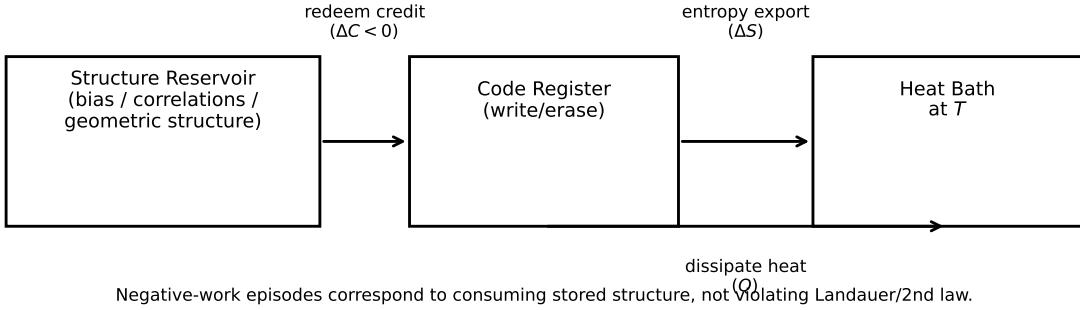


Figure 1: **Credit ledger view.** Structured reservoirs (bias, correlations, geometric structure) act as an information credit that can be redeemed to reduce local work dissipation during code formation (writing/erasing). Negative-work episodes correspond to spending stored structure, not violating the second law.

2 Landauer Cost Depends on Entropy, Not Bit-Count

Consider a one-bit register $X \in \{0, 1\}$ with distribution $p(x)$. Resetting it to a standard state (e.g., $X=0$) is a logically irreversible operation. In the simplest setting, the minimal average work required satisfies

$$W_{\min} \geq k_B T \ln 2 H(X), \quad (1)$$

where $H(X)$ is the Shannon entropy in bits. A perfectly unbiased bit has $H=1$ and the familiar $k_B T \ln 2$ appears; a *tilted* bit with $p(1) = 0.1$ has $H \approx 0.469$ and can be erased for $\approx 0.469 k_B T \ln 2$.

This immediately reframes “cheap observation”: if an internal register is already biased toward the correct answer, overwriting it to record an observation removes less uncertainty and therefore dissipates less.

to-work conversion via feedback [Toyabe et al., 2010], with the general accounting formalized in feedback fluctuation theorems [Sagawa and Ueda, 2010] and stochastic thermodynamics [Seifert, 2012].

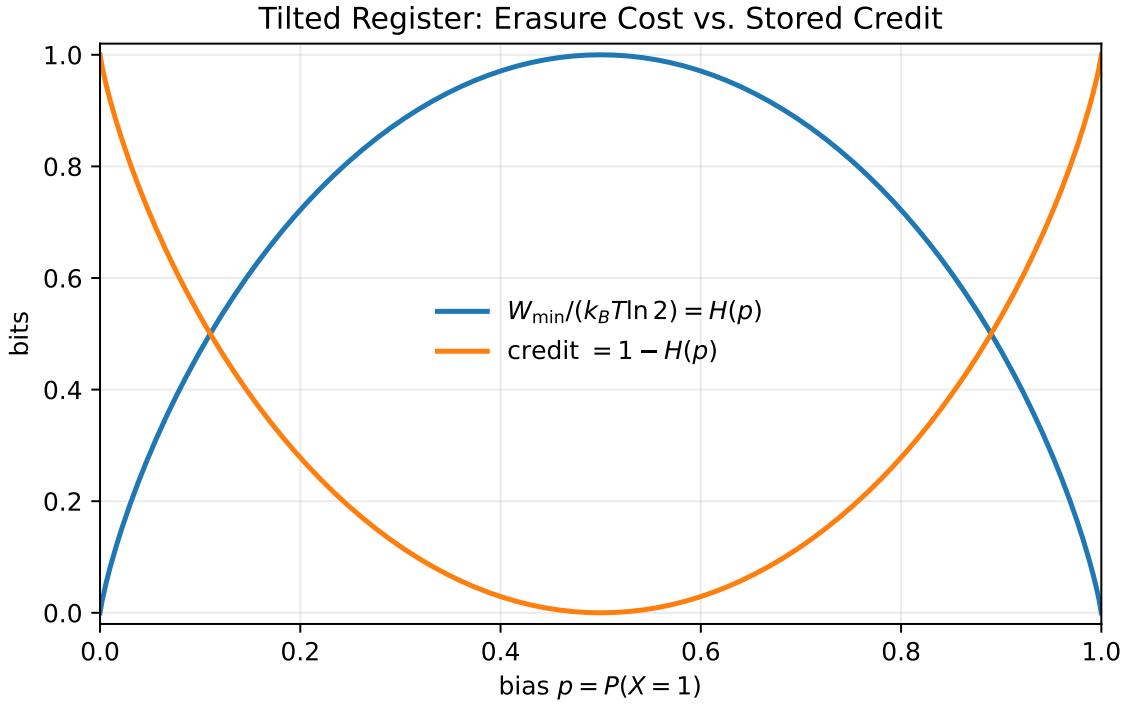


Figure 2: **Tilted register.** Minimal reset work scales with entropy removed, not with the existence of a physical bit. A pre-biased register carries “credit” (negentropy) that can be spent to make later erasure/writing cheaper.

3 Side Information: Correlation Credit

If the register X is correlated with another system Y (“environment”, “model”, “predictor”), the relevant quantity is the conditional entropy. In the presence of side information, the minimal work to reset X can be reduced to [del Rio et al., 2011, Parrondo et al., 2015]

$$W_{\min} \geq k_B T \ln 2 H(X|Y). \quad (2)$$

Using $H(X|Y) = H(X) - I(X; Y)$, the reduction in work is proportional to mutual information:

$$W_{\min} \geq k_B T \ln 2 (H(X) - I(X; Y)). \quad (3)$$

This makes the accounting explicit: if you first create mutual information (measurement, prediction, coherence), you can later *spend* it to erase or write with lower dissipation. No

paradox is required; the second law holds because building correlations consumes free energy, and consuming correlations increases entropy elsewhere.

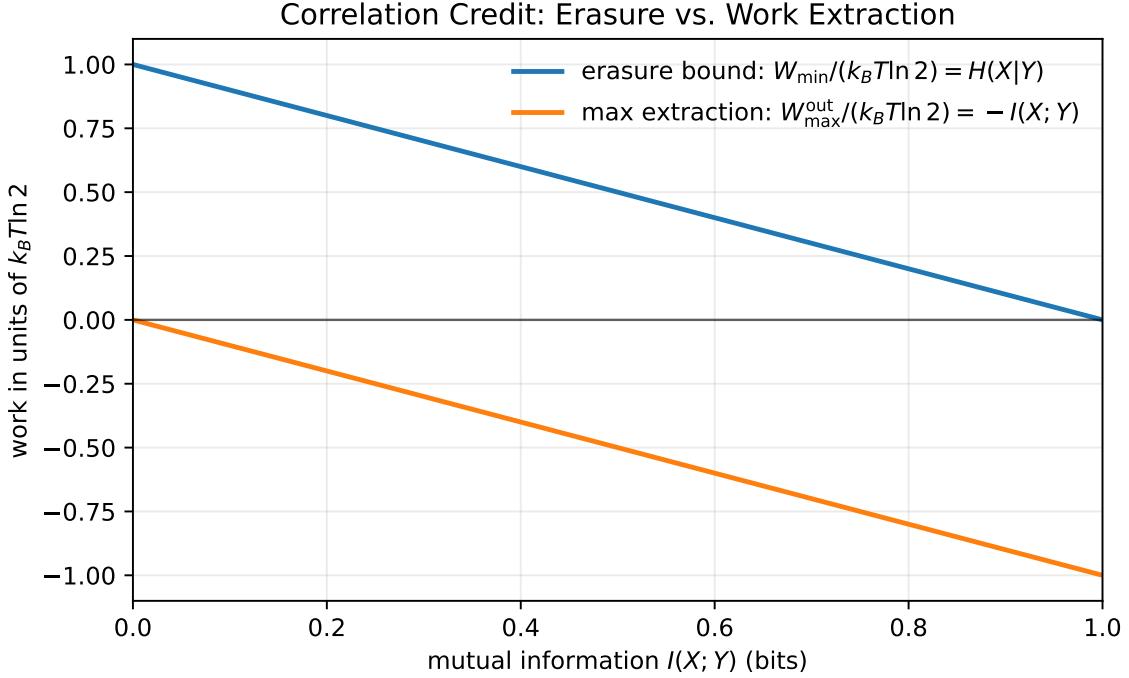


Figure 3: **Correlation credit.** Side information reduces the erasure bound from $H(X)$ to $H(X|Y) = H(X) - I(X; Y)$. For a uniform bit, $W_{\min}/(k_B T \ln 2) = 1 - I$. The same mutual information bounds maximal extractable work under feedback, giving an operational meaning to “credit” [Sagawa and Ueda, 2010, Toyabe et al., 2010].

4 Structure Credit: A Minimal Ledger

The above are standard results. The contribution here is a unified ledger interpretation that treats multiple forms of structure as a single *redeemable credit*.

4.1 Credit as nonequilibrium free energy

For a system with microstate distribution p and equilibrium distribution π (at temperature T), the nonequilibrium free energy can be written as

$$F(p) = F(\pi) + k_B T D(p\|\pi), \quad (4)$$

so the relative entropy $D(p\|\pi)$ quantifies the maximum work extractable upon relaxation. In this sense, **relative entropy is a work-credit**.

For a binary register with $\pi = (1/2, 1/2)$, $D(p\|\pi) = \ln 2 - H_{\text{nats}}(p)$, i.e., the “credit” is exactly the bit’s negentropy.

4.2 Geometric structure and thermodynamic length

Bias and correlation are not the only sources of reduced dissipation. Even when initial and final states are fixed, the energetic cost of *how* you move through state space depends on the control geometry. In linear response near equilibrium, the excess (irreversible) work for a protocol of duration τ satisfies [Crooks, 2007, Sivak and Crooks, 2012]

$$W_{\text{irr}} \geq \frac{k_B T}{2\tau} \mathcal{L}^2, \quad (5)$$

where $\mathcal{L} = \int \sqrt{d\lambda^\top g(\lambda) d\lambda}$ is the **thermodynamic length** of the path in control-parameter space under an appropriate metric g (often Fisher-information-derived). Geodesics minimize \mathcal{L} , hence minimize the unavoidable dissipation for fixed τ .

This is sometimes described loosely as “curvature makes some paths cheaper”, but what matters operationally is *anisotropic metric weighting*: some directions in state space are expensive, and good protocols move along cheap directions (or temporarily move to regions where expensive directions become cheap).

As a concrete illustration, the family of univariate Gaussian distributions has a Fisher

metric

$$ds^2 = \frac{d\mu^2 + 2 d\sigma^2}{\sigma^2}, \quad (6)$$

which corresponds (up to scale) to the Poincaré half-plane geometry after the change of variables $y = \sqrt{2}\sigma$. In this metric, changing the mean μ is especially expensive when σ is small. Figure 4 compares an axis-aligned “naive” protocol (shift μ at small σ) to the Fisher-geodesic between the same endpoints (inflate σ , move μ , then contract), illustrating how geometry selects low-dissipation paths.

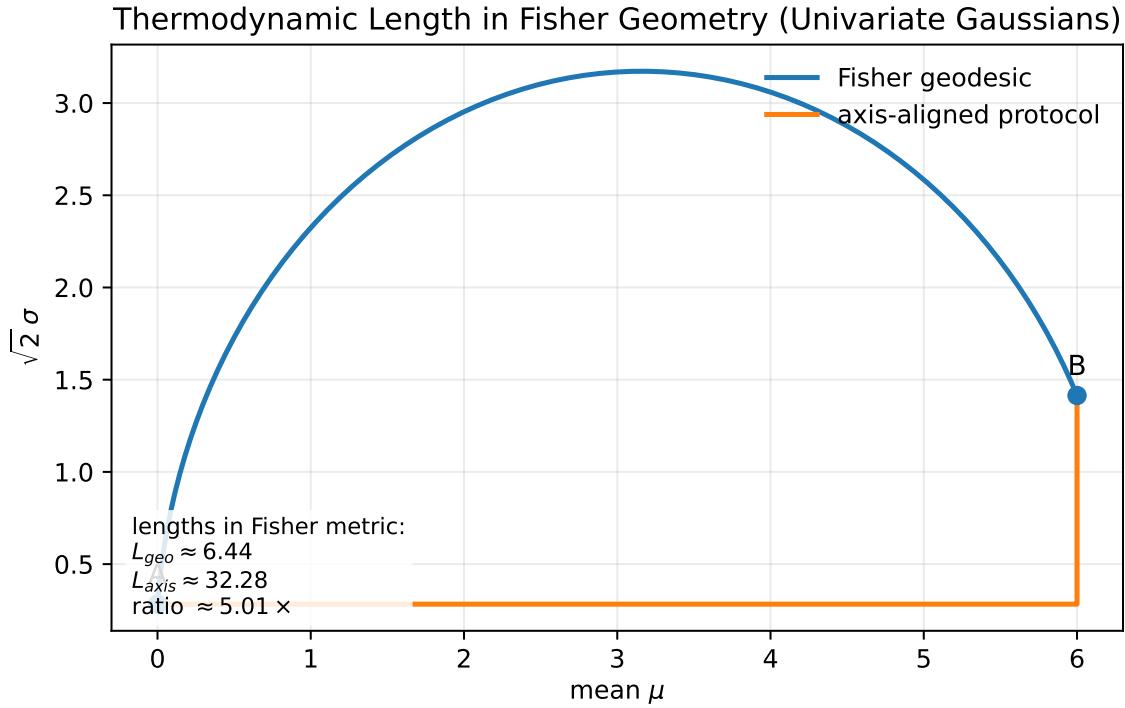


Figure 4: **Geometry constrains cost.** In the Fisher geometry of univariate Gaussians, a geodesic is shorter than a naive axis-aligned protocol between the same endpoints. In finite-time control near equilibrium, $W_{\text{irr}} \gtrsim (k_B T / 2\tau) \mathcal{L}^2$, so reductions in \mathcal{L} translate directly into reductions of unavoidable dissipation [Crooks, 2007, Sivak and Crooks, 2012].

4.3 A credit-adjusted Landauer inequality

We consider code formation steps that reduce a register's entropy by ΔH bits (e.g., from an uncertain internal state to a discrete symbol). A useful separation is:

$$W = W_{\text{rev}} + W_{\text{irr}}, \quad (7)$$

where W_{rev} is the reversible endpoint cost (entropy reduction financed by available structure), and W_{irr} is the finite-time excess dissipation governed by protocol geometry (Eq. 5).

As a coarse ledger, we track two distinct resources:

- **Correlation credit** via mutual information ΔI with side information (Section 3).
- **Structure credit** ΔC_{str} capturing other redeemable nonequilibrium structure (e.g., biased priors, chemical gradients, coherent order parameters), defined to *exclude* mutual information to avoid double counting.

In these terms, one convenient reversible bound (not claimed as a new theorem, but a useful bookkeeping identity) is:

$$\beta W_{\text{rev}} \gtrsim \Delta H \ln 2 - \Delta I - \Delta C_{\text{str}}, \quad (8)$$

where $\beta = (k_B T)^{-1}$ and ΔC_{str} is the redeemed structure credit (decrease in available nonequilibrium free energy / bias along the operation). If $\Delta I + \Delta C_{\text{str}} \geq \Delta H \ln 2$, the reversible work requirement can approach zero; if it exceeds it, the operation can exhibit $W_{\text{rev}} < 0$ (net work extraction) while consuming stored structure, consistent with standard information-to-work conversion results [Toyabe et al., 2010, Sagawa and Ueda, 2010, Seifert, 2012].

For finite-time protocols, the total work also includes W_{irr} , which is nonnegative and protocol-dependent via Eq. 5. Equivalently, relative to a naive protocol with length $\mathcal{L}_{\text{naive}}$, an optimized protocol yields a *geometric discount* in irreversible work of order $(k_B T / 2\tau)(\mathcal{L}_{\text{naive}}^2 - \mathcal{L}_{\text{opt}}^2)$.

Second-law sanity check. Credit cannot be created from nothing: building bias, correlations, or other nonequilibrium structure requires work input and/or entropy export elsewhere [Seifert, 2012, Parrondo et al., 2015]. The ledger is meant to make explicit *where* the payment happens, not to evade it.

The conceptual punchline is simple: **negative work does not invalidate Landauer; it indicates that the operation is financed by a structured reservoir.**

5 From Nonergodicity to Codes to Memory (Scope)

The credit ledger above is local: it accounts for the thermodynamic “financing” of specific code-formation steps. Here we briefly situate it in a broader dynamical ordering relevant to biological systems, where persistent bias can arise without explicit symbolic memory.

The proposed sequence is:

1. **Nonergodicity first:** High-dimensional dissipative dynamics generically confine trajectories and create history dependence (“tilt”) without explicit symbols.
2. **Code formation next:** Once confinement becomes addressable by low-dimensional control parameters (entrainment, bottlenecks, coarse-graining), stable codes emerge as handles on high-dimensional state.
3. **Memory last:** Codes become memory when stabilized on slow variables (weights, marks, morphology), making the tilt persistent across time.

In this view, “intelligence” begins as cheap nonergodic bias, and explicit symbolic memory appears later as a higher-cost but higher-portability regime.

6 Predictions and Minimal Tests

The credit ledger suggests measurable signatures that distinguish “cheap” from “forced” code formation:

6.1 Biased erasure scales with Shannon entropy

In tilted/biased erasure protocols (e.g., colloidal double-well setups), dissipation should scale as $Q \approx k_B T \ln 2 \cdot H(p)$, where p is the initial bias [Bérut et al., 2012]. “Beating” $k_B T \ln 2$ is then equivalent to preparing $H(p) < 1$.

6.2 Partial information sets the work-extraction ceiling

For feedback engines and measurement-assisted work extraction, the ceiling should track mutual information: $W_{\text{out}} \leq k_B T \ln 2 \cdot I(X; Y)$, degrading smoothly as measurement noise reduces I [Toyabe et al., 2010, Sagawa and Ueda, 2010].

6.3 Protocol geometry predicts dissipation ratios

When endpoint states are held fixed but protocols differ, dissipation should scale with thermodynamic length: $W_{\text{irr}} \propto \mathcal{L}^2 / \tau$ in the linear-response regime [Crooks, 2007, Sivak and Crooks, 2012]. This predicts quantitative ratios between naive and optimized protocols (Fig. 4).

6.4 Credit depletion implies hysteresis in finite reservoirs

If the “credit” resides in a finite, relaxational reservoir (e.g., a biased prior that is not actively re-established, a finite chemical gradient, or a coherence order parameter that decays), repeated code-formation operations should show hysteresis: early operations are cheap, later ones approach the unbiased/unstructured cost unless credit is replenished. Distinguishing “cheap because structured” from “cheap because slow” requires tracking both dissipation and recovery dynamics across repeated cycles.

7 Limitations

The credit ledger is an accounting perspective, not a full dynamical theory. It does not by itself predict (i) how quickly credit can be accrued, (ii) how different forms of credit interconvert in far-from-equilibrium regimes, or (iii) when linear-response geometric bounds (Eq. 5) are accurate. Those require specifying concrete physical substrates and protocols.

Dimension-relative information. Entropy, mutual information, and “surprise” are defined only relative to a chosen description: a set of state variables, a measurement channel, and an implicit partition/coarse-graining of state space. A “bit” presupposes that a binary distinction is representable and distinguishable in that description. Under projection or dimensional collapse, distinct high-dimensional states can alias to the same low-dimensional observation, making some distinctions (and hence some bits) undefined at that level of analysis. Figure 5 gives a toy illustration: the “bit” distinguishing two clusters does not exist at $k=1$ but becomes representable once the embedding captures the separating dimensions. More generally, injective embeddings (and faithful reconstructions) impose dimensional prerequisites [Whitney, 1936, Takens, 1981, Sauer et al., 1991]. The present ledger assumes a fixed state manifold and a fixed representational frame; extending it to variable-dimensional codes would require tracking how the admissible state space (and its metric/entropy structure) changes under dimensional reduction.

8 Conclusion

Landauer bounds are routinely invoked as fixed per-bit costs, but the physically relevant quantity is the entropy removed *given available structure*. Treating bias, correlations, and geometric/protocol structure as an information credit clarifies why biological and analog substrates can exhibit sub-Landauer dissipation and even negative-work episodes while remaining fully consistent with the second law. The remaining empirical challenge is to measure where credit is stored (bias, correlations, nonequilibrium free energy) and how rapidly

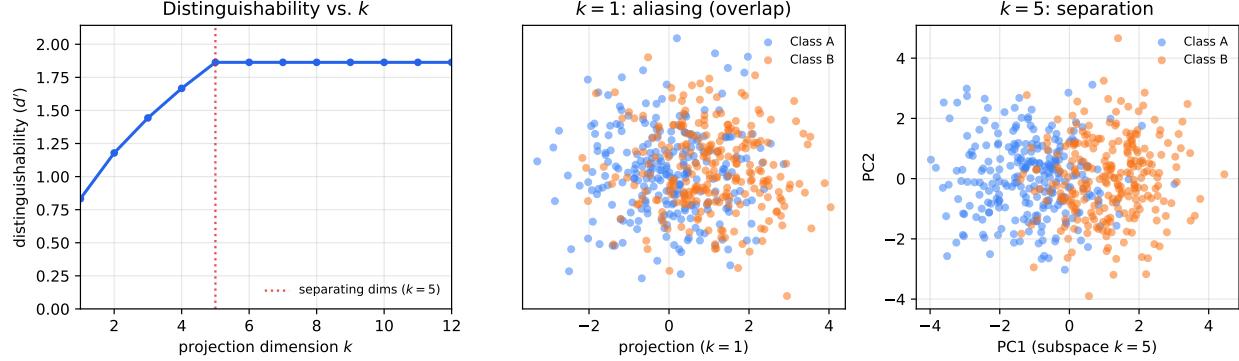


Figure 5: **Bits depend on embedding dimension.** Two clusters separated in 5 of 20 dimensions. Left: a distinguishability measure increases with projection dimension until all separating dimensions are included. Center: at $k=1$, the clusters overlap (alias) and the distinction is unrepresentable at this resolution. Right: at $k=5$, the clusters separate and the distinction becomes representable.

it can be redeemed under real dynamical constraints.

References

- Charles H. Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.
- Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider, and Eric Lutz. Experimental verification of Landauer’s principle linking information and thermodynamics. *Nature*, 483:187–189, 2012.
- Gavin E. Crooks. Measuring thermodynamic length. *Physical Review Letters*, 99:100602, 2007.
- L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474:61–63, 2011.
- R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. Thermodynamics of information. *Nature Physics*, 11:131–139, 2015.

Takahiro Sagawa and Masahito Ueda. Generalized Jarzynski equality under nonequilibrium feedback control. *Physical Review Letters*, 104:090602, 2010.

Timothy Sauer, James A. Yorke, and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65(3–4):579–616, 1991.

Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75:126001, 2012.

David A. Sivak and Gavin E. Crooks. Thermodynamic metrics and optimal paths. *Physical Review Letters*, 108:190602, 2012.

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer, 1981.

Shoichi Toyabe, Takahiro Sagawa, Masahito Ueda, Eiro Muneyuki, and Masaki Sano. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nature Physics*, 6:988–992, 2010.

Hassler Whitney. Differentiable manifolds. *Annals of Mathematics*, 37(3):645–680, 1936.