Driven Binary Memory with Failure Detection and Fuelled Repair

Brian Theory
AI-assisted research draft

November 13, 2026

D10 — Revised technical note with an explicit maintenance network. The physical model is a specialization of established information thermodynamics. The closed forms and design optimization are elementary derivations, not a claim of a new universal thermodynamic law. Internal review is not external peer review.

Abstract

What does it cost to keep a finite physical memory informative about a changing signal? We analyze a binary memory driven by a symmetric two-state signal, with a specified energy gap, thermal transition rates, and externally supplied work. The stationary tracking error, future prediction risk, work input, heat output and information-flow rate are derived explicitly. With a fixed maximum response rate, we obtain the exact minimum stationary power within this two-parameter device family for a prescribed error tolerance. A finite energy-gap restoration protocol prices a scheduled change. We then add physical health and detector states, reversible repair reactions and a maintained chemical fuel source. A symmetry theorem shows why one memory snapshot cannot diagnose health, and a two-channel counterexample exposes dissipation hidden by coarse repair rates. The baseline formulas are checked against the full four-state master equation and numerical integration; the extension is checked with independently assembled sixteen-state and eight-state generators. This model does not identify a universal price of learning, include all costs of the signal generator or controller, or solve unknown-reference recalibration. The extended maintenance subsystem has autonomous internal kinetics, but finite chemical fuel, microscopic hardware validation and optimal feedback remain open; an explicit finite battery and a common-budget controller menu are analyzed separately. The result is a small physically grounded bridge from the portfolio's abstract resource contracts to an implementation-specific design calculation.

1 Question and relation to established work

D01 separates target uncertainty from memory entropy; D05 assigns resources to maintaining diagnostic capacity. Neither accounting system alone identifies physical work or heat. Existing MIND constructions explicitly separate Boolean operators, stochastic channels and physical implementations [1, Section 1]. The present question is narrower: for a declared thermal memory and externally changing signal, what stationary power realizes a given prediction capability?

The relevant baseline is mature. Horowitz and Esposito formulate bipartite Markov thermodynamics and split mutual-information flow across subsystems [8, Sections II–III, equations (1)–(12)]. Barato, Hartich and Seifert analyze cellular sensing, including a four-state model and bounds connecting its learning rate and entropy production [3, Sections 2–4]. More precisely, the four-state generator is exactly the toy model of Barato, Hartich and Seifert [2, Section II, Fig. 1, equations

(3)–(4)] under $k_+ = k(1-\varepsilon)$ and $k_- = k\varepsilon$. Their entropy-production formula becomes (10). The present work does not introduce that generator or dissipation formula. We retain its rates in a response-budget parametrization and solve a constrained design problem; priority for that elementary optimization is not established.

Still and colleagues consider a system driven without feedback, with alternating signal-change and relaxation steps; their dissipation–prediction identity uses a specific nonequilibrium free-energy convention [13, equations (1), (8), (14)–(18)]. We do not substitute our stationary continuous-time quantities into that discrete-step identity. Instead all balances below are reconstructed directly. Kolchinsky and Wolpert define semantic information through explicit scrambling interventions and a chosen viability functional [9, Sections 3–5]. Our prediction risk is not their viability functional, and no semantic-information or autonomy theorem is inferred merely from a correlation.

2 Physical and operational contract

2.1 States, rates, and energy bookkeeping

The external signal $X_t \in \{0,1\}$ flips at rate $\gamma > 0$, independently of the memory. The physical memory $M_t \in \{0,1\}$ has energy

$$E(m,x) = \Delta\, \mathbf{1}\{m \neq x\}, \qquad \Delta \geq 0. \tag{1}$$

It is coupled to a single bath at temperature $T > 0$. Set $\beta = (k_{\mathrm{B}}T)^{-1}$ and

$$\varepsilon = \frac{1}{1+e^{\beta\Delta}} \in (0, 1/2], \qquad \begin{cases} (x,m) \longrightarrow (x, 1-m) \text{ at rate } k(1-\varepsilon), & m \neq x, \\ (x,m) \longrightarrow (x, 1-m) \text{ at rate } k\varepsilon, & m = x, \end{cases} \tag{2}$$

where $k > 0$ is the sum of the two conditional response rates. Signal transitions are $(x,m) \to (1-x,m)$ at rate γ. Simultaneous jumps have probability zero. Time is continuous; rates have units of inverse time.

The memory transition-rate ratio is $e^{\beta\Delta}$, so memory jumps satisfy local detailed balance for the stated energy gap. A signal jump changes the energy landscape. Its energy change is supplied as work by an external modulation mechanism, not as heat from the memory bath. Thus one must *not* interpret both coordinate flips as passive thermal jumps with respect to E. If all four transitions were passive in one bath, the signal rates would generally depend on m and the contract would change.

Let Q_{out} denote heat delivered to the bath, positive when the memory loses energy in a thermal jump; W denotes work supplied by signal modulation or a scheduled change of Δ. The first law is

$$\Delta U = W - Q_{\mathrm{out}}. \tag{3}$$

This boundary includes the memory's energy and the work delivered to it. The signal generator, clock, feedback electronics, fabrication, sensing readout and controller are not given microscopic dynamics. Their additional costs are outside the reported power. An ideal externally driven experiment is a permissible model, but is not a self-maintaining autonomous organism.

2.2 Information, policies, and target

The observer may read the current M_t at a declared time and predict $X_{t+\tau}$ for a fixed horizon $\tau \geq 0$ without future readouts. Only the current memory is available, not the whole memory trajectory or

the signal itself. For this experiment, readout is ideal and its apparatus cost is excluded as above. The loss is prediction error probability. Parameters (γ, k, ε) are known; no estimation phase or unknown sensor orientation is hidden in the calculation.

In the design problem, the admissible devices have one fixed $k \in (0, K]$ and one fixed finite $\Delta \geq 0$ over the evaluation interval. They are compared in stationarity under the same γ, T, K, τ, same readout rights and same target. Adaptive driving, extra memory states, time-dependent steady protocols, trajectory decoding and different reservoirs are outside this policy class. The finite restoration protocol later is a separate nonstationary experiment, not a competitor smuggled into this optimum.

Information is measured in *nats* here. Multiplication by k_{B} converts a Shannon entropy of physical states to the corresponding coarse-grained entropy contribution; multiplication by $k_{\mathrm{B}}T$ converts a dimensionless free-energy expression to energy under the stated physical assumptions. Internal degeneracies of either logical state are assumed equal and absorbed into Δ. Different degeneracies would change the free-energy gap and rates.

3 Stationary tracking and future prediction

Let $q_t = \Pr(M_t \neq X_t)$. Signal flips exchange match and mismatch. Memory transitions create mismatch at rate $k\varepsilon$ from a match and remove it at rate $k(1-\varepsilon)$ from a mismatch. Consequently

$$\dot{q}_t = \gamma(1 - 2q_t) + k\varepsilon(1 - q_t) - k(1-\varepsilon)q_t = \gamma + k\varepsilon - (2\gamma + k)q_t. \tag{4}$$

Proposition 3.1 (Stationary error and response)**.** *For $\gamma, k > 0$ and $0 < \varepsilon \leq 1/2$, the four-state chain has a unique stationary law,*

$$q = \frac{\gamma + k\varepsilon}{2\gamma + k}, \qquad p(x, m) = \begin{cases} (1-q)/2, & x = m, \\ q/2, & x \neq m. \end{cases} \tag{5}$$

For an arbitrary initial mismatch probability,

$$q_t = q + (q_0 - q)e^{-(2\gamma + k)t}. \tag{6}$$

Proof. All four states communicate since the signal flips and both memory directions have positive rates, so a finite irreducible chain has a unique stationary law. The generator is invariant under $(x, m) \mapsto (1 - x, 1 - m)$. Uniqueness therefore forces the two match probabilities to agree and the two mismatch probabilities to agree. Equation (4) fixes their sum to the displayed q; substitution verifies every balance equation. Solving the scalar linear differential equation gives (6) without requiring an initially symmetric joint law. □

The stationary marginal of either bit is uniform. Thus

$$I(X_t; M_t) = \ln 2 - h(q), \qquad h(q) = -q \ln q - (1-q)\ln(1-q). \tag{7}$$

Meanwhile $H(M_t) = \ln 2$ for every parameter value. Memory-state entropy alone cannot measure sensing quality in this example.

Proposition 3.2 (Prediction at a fixed horizon)**.** *Under the stationary law, the Bayes predictor from M_t for $X_{t+\tau}$ is $\widehat{X} = M_t$ (with a tie at $q = 1/2$), and its error is*

$$r_\tau = \frac{1}{2} + \left(q - \frac{1}{2}\right) e^{-2\gamma\tau}. \tag{8}$$

Proof. The number of future signal flips is Poisson with mean $\gamma\tau$, independent of the current pair. Its parity is odd with probability $s_\tau = (1 - e^{-2\gamma\tau})/2$. Conditional on the current memory, future mismatch therefore has probability $q(1 - s_\tau) + (1 - q)s_\tau$, giving the formula. This is at most $1/2$ because $q \leq 1/2$, so predicting the observed bit is Bayes optimal. No computation on the same single bit can beat its posterior risk. □

Tracking the current signal and predicting a distant signal are different tasks. Increasing τ pushes the error toward $1/2$ even with an informative current memory. None of the comparisons in the frontier below changes this horizon.

4 Work, heat, and information flow

The aggregate rate of thermally correcting a mismatch minus creating one is

$$J = k[(1 - \varepsilon)q - \varepsilon(1 - q)] = k(q - \varepsilon) = \gamma(1 - 2q). \tag{9}$$

A correction releases Δ; a thermal error absorbs it. A signal flip starting from a match raises the energy by Δ, while one starting from a mismatch lowers it by Δ. Therefore stationarity gives

$$P := \dot{W} = \dot{Q}_{\text{out}} = \Delta\gamma(1 - 2q), \qquad \sigma := \beta P = \gamma(1 - 2q) \ln \frac{1 - \varepsilon}{\varepsilon}. \tag{10}$$

Here σ has units of inverse time, or nats per unit time in the entropy balance. $k_{\text{B}}\sigma$ is the bath entropy current. It also equals the stationary Markov entropy-production rate of this effective four-state description, obtained by summing each undirected edge current times its forward/backward probability-flux log ratio. This does not assert that the omitted modulation apparatus has zero entropy production.

Proposition 4.1 (Information-flow specialization)**.** *The memory-transition contribution to the stationary mutual-information derivative is*

$$\ell_M = \gamma(1 - 2q) \ln \frac{1 - q}{q}, \qquad 0 \leq \ell_M \leq \sigma. \tag{11}$$

The signal contribution is $-\ell_M$, and the total time derivative of mutual information is zero in stationarity.

Proof. At the stationary symmetric law, $I = \ln 2 - h(q)$. Considering only memory jumps, $\dot{q}|_M = -J$, so their contribution to $\dot{I}$ is $J \ln((1 - q)/q)$. Signal jumps give $\dot{q}|_X = \gamma(1 - 2q) = J$, with the opposite contribution. Since

$$q - \varepsilon = \frac{\gamma(1 - 2\varepsilon)}{2\gamma + k} \geq 0$$

and $q \leq 1/2$, both logarithms are nonnegative and $\ln((1 - q)/q) \leq \ln((1 - \varepsilon)/\varepsilon)$. Multiplication by $J \geq 0$ proves the bound. At $\varepsilon = 1/2$ both quantities are zero. □

This is a direct specialization of established bipartite information-flow results [8, 3], not a new inequality. A learning rate, a trajectory mutual-information rate, the instantaneous mutual information, and its total time derivative are different quantities. The last is zero here even while information is continually renewed.

5 Exact stationary frontier within the declared device family

We now optimize over the physical parameters instead of reporting one favorable device. Fix $K > 0$ and an allowed current error $\bar{q} < 1/2$. Write

$$q_{\min} = \frac{\gamma}{2\gamma + K}. \tag{12}$$

This is an infimum reached only at infinite energy gap when $k = K$; it is not a finite-parameter device.

Proposition 5.1 (Family-specific minimum power)**.** *If $q_{\min} < \bar{q} < 1/2$, the minimum stationary work rate over the admissible devices satisfying $q \le \bar{q}$ is*

$$\boxed{P_{\min}(\bar{q}; \gamma, K, T) = k_{\mathrm{B}} T \, \gamma(1 - 2\bar{q}) \ln \frac{1 - \bar{q} + \gamma(1 - 2\bar{q})/K}{\bar{q} - \gamma(1 - 2\bar{q})/K}.} \tag{13}$$

It is attained at $k = K$ and $\varepsilon = \bar{q} - \gamma(1 - 2\bar{q})/K$. No finite-gap admissible device attains $q \le q_{\min}$. A tolerance $\bar{q} \ge 1/2$ can be met with $\Delta = 0$ and zero stationary power.

Proof. Solving (5) for ε at a desired error $q < 1/2$ gives

$$\varepsilon(q, k) = q - \frac{\gamma(1 - 2q)}{k}. \tag{14}$$

Positivity of ε is equivalent to $q > \gamma/(2\gamma + k)$, proving the feasibility condition and the unattained lower boundary. At fixed feasible q, increasing k increases ε and decreases $\ln((1 - \varepsilon)/\varepsilon)$. The other factor $\gamma(1 - 2q)$ is fixed, so power is minimized at K.

It remains to check that a stricter error than requested cannot lower power. At fixed K, both positive factors $1 - 2q$ and $\ln((1 - \varepsilon(q, K))/\varepsilon(q, K))$ strictly decrease as q increases in $(q_{\min}, 1/2)$. Their product therefore strictly decreases. The optimum subject to $q \le \bar{q}$ is at $q = \bar{q}$. Substitution into (10) proves the expression and its attainability by finite Δ. At $q = 1/2$, $\varepsilon = 1/2$ and $\Delta = 0$ provide the zero-power case. □

For a future error tolerance $\bar{r} < 1/2$ at fixed τ, use

$$\bar{q} = \frac{1}{2} + e^{2\gamma\tau} \left(\bar{r} - \frac{1}{2} \right)$$

in Proposition 5.1, checking feasibility first. The optimal value diverges as the required q approaches $q_{\min}$ from above. Unlimited response speed would change the admissible family: as $K \to \infty$, the frontier approaches $k_{\mathrm{B}} T \gamma(1 - 2\bar{q}) \ln((1 - \bar{q})/\bar{q})$. This is the information-flow lower envelope for this family, not evidence that an arbitrarily fast physical apparatus is free.

6 A physically accounted restoration protocol

An abstract cost for "recalibration" does not specify where energy goes. As a minimal physically priced modification, suppose a known weakened coupling with gap Δ_0 is restored at time zero to a stronger $\Delta_1 > \Delta_0$, while k, γ, T remain fixed. The change is a scheduled, instantaneous Hamiltonian quench: no state changes during the switch. The pre-quench distribution is stationary with error q_0; afterwards the target stationary error is q_1 and the new thermal parameter is ε_1.

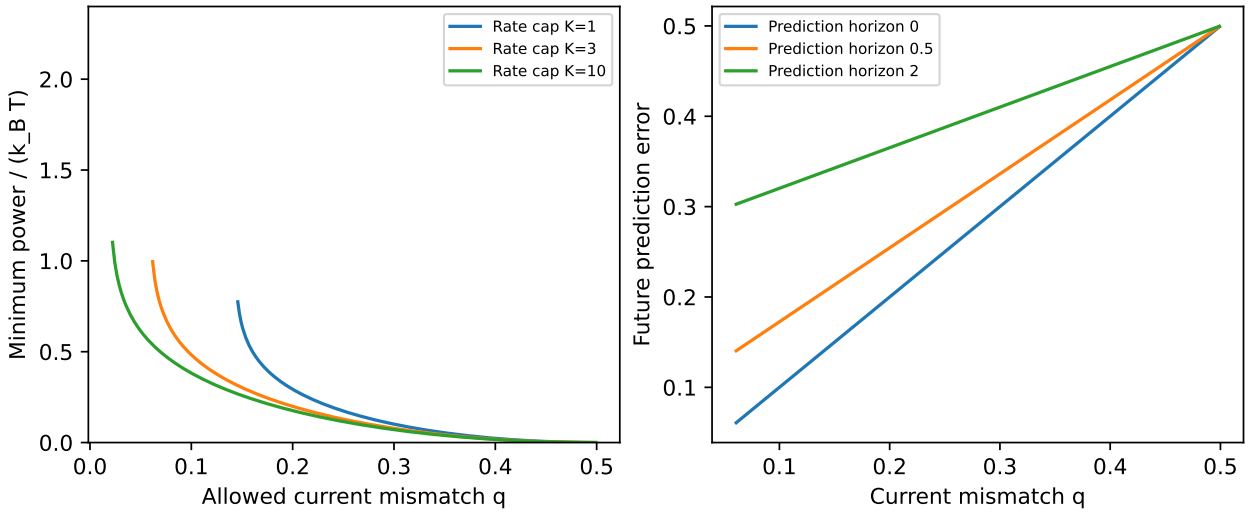


Figure 1: Analytic design curves at signal rate $\gamma = 0.2$. Left: minimum stationary power for the permitted response-rate caps; portions above the displayed vertical range are clipped for legibility, and each curve ends at its own unattainable error floor. Right: future error at fixed horizons, showing that the same current memory has different value for different prediction deadlines. Units are $k_{\mathrm{B}}T$ per time unit for power. These are model calculations, not measured device data.

This is *restoring a known coupling energy.* It is not inferring an unknown calibration, discovering a broken reference, manufacturing a new component, or autonomously deciding when to repair. The control schedule and switch mechanism are assumptions. Within those assumptions, the following accounting is exact.

Proposition 6.1 (Finite restoration balance)**.** *Let* $\lambda = 2\gamma + k$ *and*

$$B(t) = q_1 t + \frac{q_0 - q_1}{\lambda}(1 - e^{-\lambda t}).$$

Then the quench work, subsequent signal-driving work and subsequent heat output are

$$W_{\mathrm{quench}} = q_0(\Delta_1 - \Delta_0), \tag{15}$$

$$W_{\mathrm{drive}}(t) = \Delta_1 \gamma\,[t - 2B(t)], \tag{16}$$

$$Q_{\mathrm{out}}(t) = \Delta_1 k\,[B(t) - \varepsilon_1 t]. \tag{17}$$

They satisfy

$$\Delta_1 q_t - \Delta_0 q_0 = W_{\mathrm{quench}} + W_{\mathrm{drive}}(t) - Q_{\mathrm{out}}(t), \qquad q_t = q_1 + (q_0 - q_1)e^{-\lambda t}. \tag{18}$$

Proof. The quench changes energy only on the mismatch states, of total probability q_0, giving the first expression. At later times, the work rate is $\Delta_1\gamma(1 - 2q_t)$ and the heat rate is $\Delta_1 k(q_t - \varepsilon_1)$, by the same state-transition accounting as before. Since $\int_0^t q_s\,ds = B(t)$, integration gives the displayed formulas. Subtracting the heat rate from the work rate gives $\Delta_1 \dot{q}_t$ by (4); integrating and adding the quench establishes the first law. □

For $\gamma = 0.2$, $k = 3$, $\varepsilon_0 = 0.3$ and $\varepsilon_1 = 0.05$ in units $k_{\mathrm{B}}T = 1$, the mismatch changes from $q_0 \simeq 0.323529$ toward $q_1 \simeq 0.102941$. The quench supplies about 0.678487 energy units. At time two, total supplied work is about 1.537452, heat output about 1.507750, and the difference remains stored energy. These values come from the included computation and obey the finite balance, not only a stationary equality.

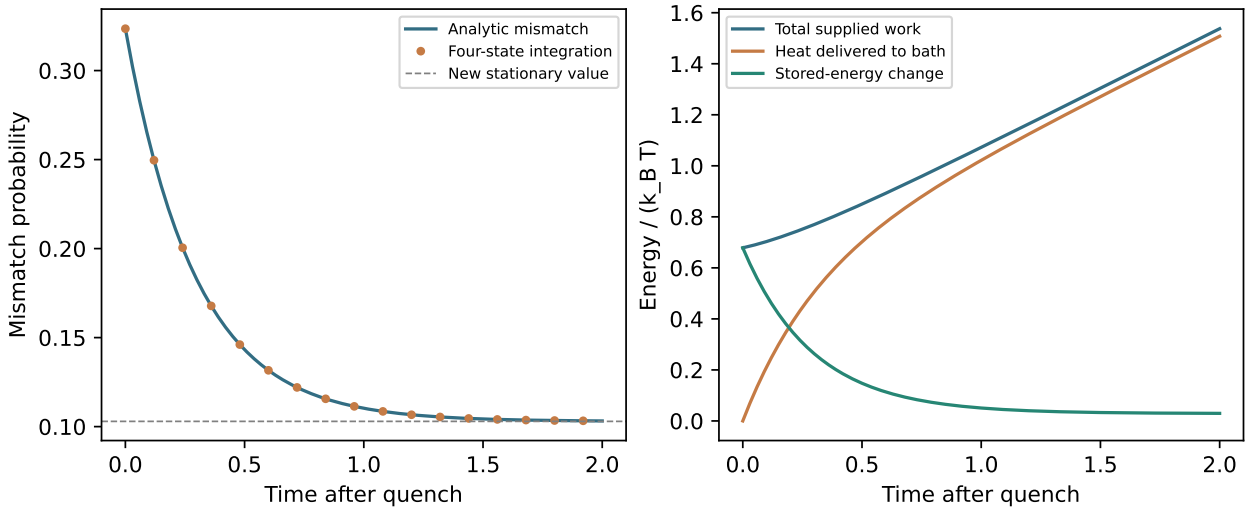


Figure 2: Specified energy-gap restoration. The analytic mismatch trajectory agrees with numerical integration of the full four-state master equation. The right panel includes the initial quench work; omitting it would give an incorrect energy budget. The energy difference is the changing stored energy. Parameter values are stated in Section 6.

7 A physical detector and fuelled repair network

The four-state frontier concerns a fixed device, not maintenance. We now give a separate, time-homogeneous finite-state model in which failure, detection and repair are transitions. Its thermodynamic laws are standard channel-resolved Markov accounting [5, Section II, equations (3)–(19)]; the contribution is a specified, reproducible maintenance example with explicit limitations.

7.1 States and local detailed balance

Let $z = (x, m, h, c) \in \{0,1\}^4$. The health state $h = 0$ denotes a weak memory and $h = 1$ a strong memory; $c = 0$ is a physical alarm state. Define

$$E(z) = \Delta_h \mathbf{1}\{x \neq m\} + \eta \mathbf{1}\{c \neq h\}, \quad 0 \leq \Delta_0 < \Delta_1, \quad \eta \geq 0. \tag{19}$$

All states have equal internal degeneracy. There is a common bath at T, a maintained chemical free-energy difference $\mu \geq 0$, and the original external signal. For a coordinate flip $z \to z'$, put $dE = E(z') - E(z)$. Choose $k_0, k_1, \nu, a > 0$ and $b \geq 0$. The directed rates are

$$\begin{aligned}
w_X(z,z') &= \gamma && (x \text{ flip}),\\
w_M(z,z') &= \frac{k_h}{1+e^{\beta dE}} && (m \text{ flip}),\\
w_C(z,z') &= \frac{\nu}{1+e^{\beta dE}} && (c \text{ flip}),\\
w_H^{\mathrm{th}}(z,z') &= a e^{-\beta dE/2} && (h \text{ flip}),\\
w_H^{\mathrm{f}}(z,z') &= b g(c) e^{\beta[\mu(1-2h)-dE]/2} && (h \text{ flip}).
\end{aligned} \tag{20}$$

The alarm gate is $g(c) = \mathbf{1}\{c = 0\}$. On a health edge it multiplies both reaction directions, since c is unchanged. When the gate is closed, both fuel rates vanish; these absent edges contribute nothing

to entropy production. Thermal health transitions remain possible. The controller is the physical detector conformation and its assumed catalytic gating; no observer supplies h as an error-free input. No scheduled clock or digital reset is required by this model.

A fuel-channel $0 \to 1$ health transition consumes chemical free energy μ; its reverse returns μ. Thermal jumps receive no work. Every non-signal channel satisfies

$$\log \frac{w^r_{zz'}}{w^r_{z'z}} = \beta(\omega^r_{zz'} - dE), \tag{21}$$

where $\omega^r_{zz'} = \mu(1-2h)$ on fuel edges and zero otherwise. Here $\omega^r_{zz'}$ denotes energy per event, not a transition rate. On signal edges $\omega^X_{zz'} = dE$ is external work, the forward and reverse rates are equal, and no bath heat is assigned. Thus signal symmetry is not being asserted as passive thermal detailed balance.

The health transition energy is

$$dE_H = (\Delta_{1-h} - \Delta_h)\mathbf{1}\{x \neq m\} + \eta(\mathbf{1}\{c \neq 1-h\} - \mathbf{1}\{c \neq h\}). \tag{22}$$

The detector therefore acts back on health. Omitting its interaction energy from repair would violate the declared local detailed balance. Likewise a constant repair/reverse ratio cannot implement a fixed μ when dE_H varies. The reverse chemical reaction cannot simply be deleted at finite affinity. Here "repair" means conversion to the designated strong conformation, not perfect restoration or replacement of a damaged material.

This is a mesoscopic thermodynamic realization, not a fabricated controller or atomistic derivation. Attempt rates and catalytic barriers are assumed fixed hardware. The maintenance subsystem is autonomous in the sense of time-independent internal kinetics, while the whole device remains externally powered: the signal source, maintenance of the chemical reservoirs, construction, readout apparatus and eventual replacement of the controller itself remain outside the boundary. A maintained μ is not a finite fuel tank.

7.2 Channel-resolved energy and entropy balances

Let π be the unique stationary law; positivity of the thermal coordinate rates and $\gamma > 0$ gives irreducibility even if $b = 0$. Sum each unordered state pair once for each available reaction channel r, and orient it arbitrarily. Write

$$j^r_{zz'} = \pi_z w^r_{zz'} - \pi_{z'} w^r_{z'z}, \qquad J_{\mathrm{f}} = \sum_{x,m,c} j^{\mathrm{f}}_{(x,m,0,c),(x,m,1,c)}.$$

Absent fuel pairs have current zero. Define

$$\begin{aligned} P_X &= \sum_z \pi_z \gamma \Delta_h (1 - 2\mathbf{1}\{x \neq m\}), \\ P_{\mathrm{f}} &= \mu J_{\mathrm{f}}, \qquad \dot{Q}_r = \sum_{\{z,z'\}} j^r_{zz'}(\omega^r_{zz'} - dE), \\ \sigma_{\mathrm{ch}} &= \sum_{r,\{z,z'\}} j^r_{zz'} \log \frac{\pi_z w^r_{zz'}}{\pi_{z'} w^r_{z'z}}. \end{aligned} \tag{23}$$

For signal edges the heat term is zero. Signs of individual channel heat flows need not be positive.

Proposition 7.1 (Stationary maintenance balance)**.** *For the finite network above,*

$$P_X + P_{\mathrm{f}} = \sum_r \dot{Q}_r, \qquad \sigma_{\mathrm{ch}} = \beta(P_X + P_{\mathrm{f}}) \geq 0. \tag{24}$$

Consequently the fuel current and signal work, not an abstract count of detected failures, determine the stationary energy bill within this boundary.

Proof. Each transition obeys $dE = \omega^r_{zz'} - q^r_{zz'}$. Summing event rates gives the first law; stationarity sets $d\langle E\rangle/dt = 0$. In σ_{ch}, the terms $j^r_{zz'} \log(\pi_z/\pi_{z'})$ sum to the stationary Shannon entropy derivative, zero. The remaining log-rate terms equal $\beta q^r_{zz'}$ by (21); they vanish on the symmetric signal edges. This proves the equality. Each summand $(f - b)\log(f/b)$ for positive forward and backward probability fluxes is nonnegative. The assertion follows without an information-to-energy conversion postulate. □

7.3 What the memory and alarm can diagnose

A predictor is not automatically a fault detector. The following is an elementary symmetry consequence, related conceptually to D03's observation ambiguities but using a different interface: one physical snapshot rather than a fresh-target observation history.

Proposition 7.2 (Snapshot blindness)**.** *For any irreducible generator on (X, M, H, C) invariant under $(x, m, h, c) \mapsto (1 - x, 1 - m, h, c)$, its stationary law satisfies*

$$\Pr(M = m, H = h, C = c) = \tfrac{1}{2}\Pr(H = h, C = c).$$

In particular $I(M; H, C) = 0$. The minimum probability of incorrectly classifying health from M alone is $\min\{\Pr(H = 0), \Pr(H = 1)\}$, identical to using no observation. All rates in (20) *obey this symmetry.*

Proof. Uniqueness forces the stationary law to be invariant under the permutation. Summing over x gives equal probabilities for $M = 0$ and $M = 1$ at every fixed (h, c), proving independence. Every conditional health posterior given M therefore equals its prior; a majority-prior decision minimizes zero-one loss. □

The proposition does not cover a memory trajectory, a simultaneous readout of X and M, or a nonstationary initialization. Waiting times and disagreements can reveal health. Instantaneous and trajectory information are different quantities [7, Section II]. For a single alarm readout the Bayes health risk instead is

$$R_H(C) = \sum_c \min\{\Pr(H = 0, C = c), \Pr(H = 1, C = c)\}.$$

Positive $I(H; C)$ need not strictly reduce this particular risk: posteriors can move without crossing the decision threshold. The gate $g(C)$ is a specified catalytic response, not a claimed Bayes-optimal maintenance policy.

For prediction using only M, symmetry gives a symmetric (X, M) marginal. Write its mismatch probability as q; it is obtained from the full joint stationary law, not by substituting an averaged k into Proposition 3.1. Since future signal flips remain independent, the Bayes error at horizon τ is

$$r^*_\tau = \tfrac{1}{2} - |q - \tfrac{1}{2}|e^{-2\gamma\tau}. \tag{25}$$

This follows by choosing M or its complement according to the sign of $1/2 - q$. It does not require a general claim that every feedback design has $q \leq 1/2$.

7.4 A finite calculation and an adverse detector example

In units $k_{\mathrm{B}}T = 1$, take $\gamma = .2$, $(\Delta_0, \Delta_1) = (.3, 3)$, $(k_0, k_1) = (.1, 3)$, $\eta = 2$, $\nu = 2$, $a = .03$, $b = .2$, $\mu = 4$. The full 16-state generator gives

$$\begin{aligned} q &= 0.1486766841, & \Pr(H = 1) &= 0.8854655386, \\ P_X &= 0.4247794641, & P_{\mathrm{f}} &= 0.0838286267, \\ \sum_r \dot{Q}_r &= 0.5086080907, & \sigma_{\mathrm{ch}} &= 0.5086080907. \end{aligned}$$

The detector-channel heat output is 0.0400126629; this is a channel flow, not a separately identified minimum detection cost. Aggregating parallel health reactions before calculating entropy production gives only 0.4575317027, missing dissipation.

The joint law with rows $h = 0, 1$ and columns $c = 0, 1$ is

$$\Pr(H, C) \simeq \begin{pmatrix} 0.0958800361 & 0.0186544253 \\ 0.1105516624 & 0.7749138762 \end{pmatrix}.$$

Although $I(H; C) = 0.1249301762$ nats, the Bayes classifier always reports healthy, so $R_H(C) = 0.1145344614$, equal to the prior-only risk. Reporting $H = C$ has larger error 0.1292060878. The false-alarm probability given healthy is 0.1248514568, and the missed-alarm probability given weak is 0.1628717253. This example prevents equating correlation, detection accuracy and useful control.

In the recorded sensitivity grid, changing only η to 4 at $\mu = 4$ gives $R_H(C) = 0.0219343248$, below the prior-only risk 0.1112279540; the detector can be informative for classification. This is a different physical device with altered interaction and backaction, not a free improvement. The code also evaluates disabling the fuel channel and opening it for both detector states. Those settings trade prediction against work; they are neither a global optimization nor comparisons at matched risk. No advantage over the original frontier is claimed because the state space and rate contract have changed.

8 Why observed repair rates do not determine dissipation

Retaining reaction channels is essential, not merely a numerical convention. The following explicit calculation is a standard cycle-accounting specialization [5].

Proposition 8.1 (A hidden dissipative repair cycle)**.** *Let a health bit have energies* $0, E$*, where* $E \geq 0$*, with thermal and chemical channel rates*

$$a_+ = ae^{-\beta E/2}, \quad a_- = ae^{\beta E/2}, \qquad b_+ = be^{\beta(\mu - E)/2}, \quad b_- = be^{-\beta(\mu - E)/2},$$

where $a, b, \mu > 0$*. Set* $S = a_+ + a_- + b_+ + b_-$*. The stationary fuel current and entropy production are*

$$J_{\mathrm{f}} = \frac{b_+ a_- - b_- a_+}{S} = \frac{2ab \sinh(\beta\mu/2)}{S} > 0, \qquad \sigma_{\mathrm{ch}} = \beta\mu J_{\mathrm{f}} > 0. \tag{26}$$

The two-state generator after adding the two channels has zero stationary entropy production when evaluated as a single-channel process.

Proof. The stationary probabilities are $p_0 = (a_- + b_-)/S$ and $p_1 = (a_+ + b_+)/S$. Their substitution gives (26). The thermal current is $-J_{\mathrm{f}}$. Thus the two channel entropy terms combine to $J_{\mathrm{f}} \log(b_+ a_-/(b_- a_+)) = \beta\mu J_{\mathrm{f}}$. The aggregate forward and reverse probability fluxes are equal by two-state stationarity, so the aggregated expression is zero. □

For $\beta = 1, a = .3, b = .2, E = 1, \mu = 4$, $J_{\mathrm{f}} = 0.2690650022$ and $\sigma_{\mathrm{ch}} = 1.0762600086$ while the aggregate value is zero. The same aggregate rates could instead be realized by a single passive two-state system with energy difference $E_{\mathrm{eff}} = -\beta^{-1}\log[(a_+ + b_+)/(a_- + b_-)]$ and zero stationary dissipation. Its energy landscape differs. An observer with only the coarse health trajectory cannot distinguish these energetic realizations. Known energies, channel identities and reservoir couplings are additional information. This explicit counterexample strengthens the original warning that a coarse kinetic repair rate cannot identify total repair cost; it is not a new general law of hidden entropy production.

9 Finite energy storage and a matched maintenance contract

A maintained chemical potential is an inexhaustible external resource. To expose this boundary, replace the chemical channel by a finite battery $n \in \{0, \dots, N\}$ with equally degenerate levels and total energy

$$\mathcal{E}(z, n) = E(z) + \mu n.$$

This is an ideal finite energy ladder, not a finite well-mixed chemical reservoir. Finite chemical reservoirs and explicit replenishment already have substantial antecedents [6, Section II]; conservation laws in chemical thermodynamics are established [10]. A recent preprint derives open chemical-network dynamics and thermodynamics from closed systems under time-scale and abundance separation [11]; the present ladder example establishes no such limit. The construction below supplies a bounded maintenance example, not priority for finite-reservoir thermodynamics.

For $h = 0, n > 0$, pair $(x, m, 0, c, n)$ with $(x, m, 1, c, n-1)$. Write $d = E(x, m, 1, c) - E(x, m, 0, c)$ and set

$$r_+ = bg(c)e^{\beta(\mu-d)/2}, \qquad r_- = bg(c)e^{-\beta(\mu-d)/2}.$$

Both rates belong to that same pair. A forward repair consumes one level, its reverse restores one. No transition crosses a battery boundary; all other original thermal channels leave n unchanged. Signal jumps still do external work. Local detailed balance now holds against the total energy change $d - \mu$, without an external chemical-work term. The forward and reverse updates change two coordinates together, so a single-coordinate multipartite information-flow theorem must not be applied without regrouping health and battery.

Proposition 9.1 (Finite battery accounting)**.** *Let R_t^+, R_t^- count forward and reverse battery-channel events. Pathwise,*

$$R_t^+ - R_t^- = n_0 - n_t, \qquad -N \le R_t^+ - R_t^- \le N.$$

The mean energy balance for the enlarged device is

$$\Delta\langle\mathcal{E}\rangle = W_X - Q_{\mathrm{out}},$$

whereas the memory/controller boundary alone satisfies

$$\Delta\langle E\rangle = W_X + W_B - Q_{\mathrm{out}}, \qquad W_B = \mu\langle n_0 - n_t\rangle.$$

In every stationary law the net battery current is zero. Moreover, $\langle R_t^+ - R_t^-\rangle/t \to 0$ as $t \to \infty$.

Proof. Only battery-channel events change n, by respectively minus or plus one. Summing these increments gives the pathwise identity and bound. For each such event the battery-energy loss cancels work delivered to the device; summing all event energies gives the first laws. Stationarity implies $d\langle n\rangle/dt = 0$; the long-time conclusion also follows directly from boundedness, without irreducibility. □

W_B denotes net energy drawn from the ladder, not automatically reversible work or free-energy depletion. The battery distribution can gain entropy and correlate with the controller; extractable-work accounting would additionally require those terms. The budget below constrains this explicit energy statistic.

The proposition does not bound the gross number of repairs by N. A thermal health failure followed by repair discharges the battery; a reverse battery event followed by thermal recovery charges it. Thus $n = 0$ is generally not absorbing, and the externally driven signal can supply energy to the enlarged system. For $b = 0$, battery sectors are disconnected; no unique stationary law on the whole enlarged space is asserted. A finite battery cannot indefinitely provide positive *net* stationary power, but this does not prove that prediction or maintenance must cease.

9.1 What a chemical realization would additionally require

For an ideal mixture with n fuel and $N - n$ waste molecules, the multiplicity is $\binom{N}{n}$. If the energy difference per fuel molecule is μ_0, its coarse free energy is

$$F_B(n) = n\mu_0 - k_{\mathrm{B}}T \log \binom{N}{n} + \text{constant}, \quad F_B(n) - F_B(n-1) = \mu_0 + k_{\mathrm{B}}T \log \frac{n}{N-n+1}.$$

A mass-action pair would therefore have an occupancy factor $n/(N - n + 1)$ in its forward/reverse ratio, in addition to $e^{\beta(\mu_0 - d)}$. Fixed interior rates and fixed μ do not model this mixture. The ladder calculations are retained as a transparent finite-storage contract; the corresponding chemical-mixture experiment has not been performed.

Local detailed balance provides a limited realizability certificate. For any retained thermal pair with energies $\mathcal{E}_i, \mathcal{E}_j$ and positive rates satisfying $r_{ij}/r_{ji} = e^{-\beta(\mathcal{E}_j - \mathcal{E}_i)}$, choose an attempt frequency $\nu_* \geq \max(r_{ij}, r_{ji})$ and put

$$B_{ij} = \mathcal{E}_i - \beta^{-1} \log(r_{ij}/\nu_*).$$

The ratio condition gives the same B_{ij} from the reverse direction, and $B_{ij} \geq \max(\mathcal{E}_i, \mathcal{E}_j)$. Thus the pair admits an abstract common-barrier Arrhenius representation. This is algebraic consistency, not a molecular design: independently chosen barriers may not be jointly constructible. Exact closed gates correspond to absent edges or ideal infinite barriers. A physical implementation must measure leakage, prefactors, detector backaction and channel identities. The external signal edges remain externally actuated and are excluded from this passive-barrier certificate.

9.2 Comparison rules and evidence

Every comparator uses the same initial state distribution, horizon, battery capacity, energy landscape, detector, and cap on catalytic prefactor. Differences in actual energy consumed are reported, not hidden behind a common capacity. We consider a finite declared menu of constant prefactors and alarm or always-open gates; disabling the battery channel is included. Neither exhaustive feedback optimization nor a hardware-independent maintenance advantage is claimed.

The energy-budget statistic is expected net input to the memory/controller, $W_X + W_B$, including signed signal work. A bound on this expectation is not a hard trajectory budget, an equal-risk comparison, or a bound on gross external supply. Preparation and charging of the shared initial state are excluded equally. End-state battery charge and stored device energy are reported to make terminal accounting explicit. A different convention forbidding recovery of negative signal work requires a different budget analysis. With the parameters of Section 7, take $N = 4$, horizon 20, and the shared initial law assigning probability $1/2$ to each $x = m = 0$ and $x = m = 1$, with

$h = c = 1, n = 4$. The fixed prediction rule is $\widehat{X} = M$; its objective is integrated mismatch, not an optimized trajectory decoder. The menu is $b \in \{0, .025, .05, .1, .2\}$ with alarm and always-open gates. The latter opens more edges; a common per-edge prefactor cap is not an equal aggregate-activity budget. Selected results (energy in $k_{\mathrm{B}}T$) are:

Gate	b	$\int_0^{20} q_t\, dt$	$W_X + W_B$	$\langle n_{20} \rangle$
alarm	0.00	3.297951	7.862265	4.000000
alarm	0.05	2.938802	9.155016	3.810656
alarm	0.20	2.546346	10.340105	3.652453
always	0.05	2.690954	9.888171	3.714516
always	0.20	2.323822	10.871469	3.591738

Under expected-net-input budget 10, the best menu entry is always-open with $b = .05$: integrated error 2.690954, input 9.888171. At budget 15, always-open $b = .2$ gives 2.323822 and 10.871469. Budget 5 admits no menu entry, including disabling repair; this is menu infeasibility, not a universal impossibility theorem. These two winners provide no alarm-gating advantage at those budgets. At intermediate budgets 9 and 9.8, however, the best listed entries are alarm gates with $b = .025$ and $b = .1$, respectively. Thus the menu does not establish general dominance of either gate. No interpolation, continuous optimization or guarantee for unlisted controllers is inferred.

The initial device energy is zero. The final device energies of the budget-10 and budget-15 winners are 0.518543 and 0.531496, respectively; all rows retain this field in the supplied data.

The code records 18 horizon/capacity/gate evaluations ($N = 1, 4$; horizons $5, 20, 100$; alarm, always-open and off) and ten menu rows, including duplicates. It reports positive-only signal work separately, gross forward/reverse events, final empty probability and time spent empty. All local-balance, normalization and finite-first-law checks are below 10^{-9}. In the $N = 4$ stationary computations at $b = .2$, net battery power is numerically zero, yet forward event rates are 0.003725 (alarm) and 0.007469 (always-open), matched by reverse events. These gross cycles do not demonstrate useful work. Empty-state probability is about .981 but emptiness is not absorbing.

10 Counterexamples and limits of the interpretation

A kinetic slowdown has no identified repair price here. If degradation only lowers k while leaving Δ fixed, the generator describes poorer tracking. It does not specify the physical mechanism for restoring k. Raising a transition barrier, replacing a component, and changing catalytic activity can share a coarse rate but have different control costs. One can also append a dissipative actuator that leaves the displayed generator unchanged. Therefore the generator alone cannot determine the total cost of repairing that kinetic parameter. A number assigned without a microscopic protocol would be an abstract resource cost, as in D05.

State entropy does not track useful information. $H(M) = \ln 2$ throughout the stationary family. Reducing q increases correlation and raises the minimum power while leaving this marginal entropy unchanged. Thus "lower physical or memory entropy means better knowledge" is not the result.

No universal power per learned bit. The rate ℓ_M and stationary power depend on γ, k, Δ and on the boundary. At $\gamma \to 0$, a nearly static signal needs vanishing stationary refresh power even though an informative equilibrium memory can remain correlated with it. Preparation costs

and the noncommuting transient limit are separate. This does not contradict a finite cost for a prescribed reset protocol. Likewise a different reservoir or extra state can change the frontier, which was optimized only over (2).

A fixed performance target matters. Allowing a predictor to wait longer changes $X_{t+\tau}$ and its uncertainty; comparing it against an earlier target is not a fair performance comparison. Here every design uses the same horizon and current-memory-only readout. A trajectory filter could extract more information and would require a larger controller contract.

Prediction is not autonomous maintenance. The original four-state memory has no health detector. Section 7 supplies an explicit detector and repair channel, but does not repair its work source, choose a system boundary, or preserve a representation through replacement. Those require further dynamics. Accurate prediction by itself establishes none of them, and establishes no phenomenal consciousness.

11 Reproducibility and independent checks

`code/verify.py` builds the generator in the order $(0,0),(0,1),(1,0),(1,1)$. It solves the four-state stationary equations for 80 parameter triples and independently sums edgewise work, heat and probability-flux entropy production. Those numerical quantities agree with the scalar formulas to below 10^{-10} in the recorded run. It also checks the information-flow inequality.

For 24 target specifications, a bounded grid search over up to 400 response rates and 160 feasible errors per rate checks the endpoint minimizer in Proposition 5.1. This finite search is not the proof and does not establish global optimality outside the declared family. The restoration experiment integrates four populations and accumulated work and heat with relative tolerance 10^{-11} and absolute tolerance 10^{-13}. The maximum reported discrepancy from the integrated formulas is below 10^{-9}. There is no stochastic seed because no Monte Carlo sampling is used. All parameters, numerical residuals, CSV data, PDF/SVG figures and plotting source are supplied.

The new `code/controller.py` keeps thermal and fuel channels separate, solves the 16-state generator and checks stationarity, local detailed balance, both thermodynamic balances and the health-blindness identity. It records a reference device, two illustrative comparators and a 16-entry (μ,η) sensitivity grid (one grid entry repeats the reference). These are bounded floating-point checks, not a global controller optimum. An independently assembled eight-state mismatch/health/detector generator and a separate integration check are recorded with the new review evidence. Original sources, data and reviews are preserved unchanged outside this revision.

The finite-battery code `code/finite_battery.py` builds $16(N+1)$ states and integrates populations and additive observables with an augmented matrix exponential. Its 28 transient rows include duplicates; the largest first-law, local-balance or normalization residual is 7.03×10^{-12}. A separate hostile AI reviewer independently built an $8(N+1)$-state lumped chain and integrated with DOP853: the largest discrepancy across all 28 rows is 1.17×10^{-11}. Both implementations and the corrected budget-menu interpretation are preserved with the reviews. This is independent numerical reproduction within a synthetic model, not empirical validation of rates or hardware.

12 Remaining research and publication assessment

The baseline sensing generator and its entropy production are established prior art. The rate-capped frontier is an elementary design specialization, with no historical-priority claim. The maintenance extension contributes an explicit mesoscopic realization, channel-aware code, snapshot observability proof and counterexamples. Its first and second laws are applications of established stochastic thermodynamics, not novel thermodynamic principles.

A fresh primary-source search also checked sensory adaptation [12], sensory capacity [7], and a 2026 preprint on a finite-time autonomous eraser/refrigerator [4]. Their tasks differ from maintaining a degrading predictive memory; that distinction neither establishes novelty nor rules out closer antecedents. The search record specifies inspected passages and incomplete coverage. The internal AI specialist also contributed the battery computation; the hostile reviewer supplied the independent implementation. Both reviews and their role disclosures are preserved. Neither is human peer review.

The principal unresolved problems are: (i) finite chemical fuel and replenishment beyond the ideal finite battery now analyzed; (ii) a matched-risk optimization over detector coupling, response rates and admissible control mechanisms; (iii) a physical realization and fabrication/actuator model for the assumed rates; (iv) trajectory-based failure diagnosis and unknown reference orientation; and (v) maintaining the detector and repair apparatus themselves. None is solved by the finite stationary example. The current model does not supply a hardware-independent detection or repair price.

D01 remains the abstract information-accounting companion. D03 concerns reference uncertainty under its own observation contract; the snapshot symmetry result here must not be restated as a theorem about all histories. D05's observed-health viability and abstract resource threshold do not acquire energy units by multiplying by $k_{\mathrm{B}}T$. A future merger requires finite fuel states, common observations, explicit task deadlines and coupled dynamics. No other portfolio manuscript was edited. The present classification remains a reproducible technical methods note, without a priority, publication-readiness or consciousness claim.

References

[1] B-Theory research programme. The MIND: Five finite constructions, 2026. URL `https://www.b-theory.com/manuscripts/mind-five-finite-constructions-r1-2026-10-05.pdf`. First research edition with P1 and P2 addenda, October 5. AI-assisted research prepared for Brian Theory. Sections 1–4 inspected.

[2] Andre C. Barato, David Hartich, and Udo Seifert. Information-theoretic versus thermodynamic entropy production in autonomous sensory networks. *Physical Review E*, 87:042104, 2013. doi: 10.1103/PhysRevE.87.042104. URL `https://arxiv.org/abs/1212.3186`. arXiv:1212.3186v3, Section II, Fig. 1 and equations (3)–(4) inspected; exact generator correspondence stated in text.

[3] Andre C. Barato, David Hartich, and Udo Seifert. Efficiency of cellular information processing. *New Journal of Physics*, 16:103024, 2014. doi: 10.1088/1367-2630/16/10/103024. URL `https://arxiv.org/abs/1405.7241`. Published version, Sections 2–4 inspected.

[4] Wanyan Chen, Miao Chen, and Yu-Han Ma. Finite-time thermodynamics of an autonomous information machine. arXiv:2604.15953v1, 2026. URL `https://arxiv.org/abs/2604.15953`.

Preprint April 17, 2026; abstract, introduction and model setup inspected. Different two-bath eraser/refrigerator task; not used as proof of this paper's results.

[5] Massimiliano Esposito. Stochastic thermodynamics under coarse graining. *Physical Review E*, 85: 041125, 2012. doi: 10.1103/PhysRevE.85.041125. URL `https://arxiv.org/abs/1112.5410`. arXiv v1 dated December 22, 2011; Section II equations (3)–(19) inspected. HTML conversion date is not publication date.

[6] Jonas H. Fritz, Basile Nguyen, and Udo Seifert. Stochastic thermodynamics of chemical reactions coupled to finite reservoirs: A case study for the brusselator. arXiv:2003.02148v2, 2020. URL `https://arxiv.org/abs/2003.02148`. Section II finite reservoirs and external pump inspected; different task, not a validation of the present hardware.

[7] David Hartich, Andre C. Barato, and Udo Seifert. Sensory capacity: An information theoretical measure of the performance of a sensor. *Physical Review E*, 93:022116, 2016. doi: 10.1103/PhysRevE.93.022116. URL `https://arxiv.org/abs/1509.02111`. arXiv:1509.02111v2; Section II setup and instantaneous versus trajectory distinction inspected.

[8] Jordan M. Horowitz and Massimiliano Esposito. Thermodynamics with continuous information flow. *Physical Review X*, 4:031015, 2014. doi: 10.1103/PhysRevX.4.031015. URL `https://arxiv.org/abs/1402.3276`. Published version, Sections II–III; arXiv v3 dated June 30, 2014.

[9] Artemy Kolchinsky and David H. Wolpert. Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface Focus*, 8:20180041, 2018. doi: 10.1098/rsfs.2018.0041. URL `https://doi.org/10.1098/rsfs.2018.0041`. Author-hosted published PDF, Sections 3–5 inspected.

[10] Riccardo Rao and Massimiliano Esposito. Conservation laws and work fluctuation relations in chemical reaction networks. arXiv:1805.12077v2, 2019. URL `https://arxiv.org/abs/1805.12077`. Author copy dated January 2019; introduction and framework scope inspected; no full proof audit.

[11] Benedikt Remlein, Massimiliano Esposito, and Francesco Avanzini. Emergence of open chemical reaction network thermodynamics within closed systems, 2026. URL `https://arxiv.org/html/2604.14848v1`. Version 1; abstract and introduction inspected.

[12] Pablo Sartori, Leo Granger, Chiu Fan Lee, and Jordan M. Horowitz. Thermodynamic costs of information processing in sensory adaptation. *PLoS Computational Biology*, 10:e1003974, 2014. doi: 10.1371/journal.pcbi.1003974. URL `https://doi.org/10.1371/journal.pcbi.1003974`. Publisher full text; model and information-processing sections inspected, not independently proof-audited.

[13] Susanne Still, David A. Sivak, Anthony J. Bell, and Gavin E. Crooks. Thermodynamics of prediction. *Physical Review Letters*, 109:120604, 2012. doi: 10.1103/PhysRevLett.109.120604. URL `https://escholarship.org/uc/item/8gf283cq`. Author reading copy inspected, problem setup and equations (1), (8), (14)–(18); publication metadata checked against publisher.