

Thermodynamic Trade-off Frontiers for Neuromorphic Processors: A Quantum-Inspired Non-Equilibrium Model of the Energy–Speed Boundary

Thermodynamic Trade-off Frontiers for Neuromorphic Processors: A Quantum-Inspired Non-Equilibrium Model of the Energy–Speed Boundary

Abstract

Neuromorphic processors promise order-of-magnitude gains in energy efficiency, but no predictive framework currently connects device-level dissipation to system-level computational throughput. We develop a non-equilibrium thermodynamic model in which a neuromorphic processor is treated as a driven, dissipative many-body system maintained in a non-equilibrium steady state by continuous energy flux. The model combines three ingredients: a Landauer-type entropy-production floor per synaptic event, a Margolus–Levitin-style quantum speed limit recast as a classical energy–time bound on switching, and a Rayleigh dissipation potential governing relaxation of membrane and synapse state variables. Energy per synaptic event is decomposed into Landauer, non-adiabatic switching, and leakage terms. For representative analog-CMOS parameters (10^6 synapses, 100 fF synapse capacitance, 0.5 V swing, 10 pW leakage per synapse), the model predicts a minimum energy per synaptic event of $\approx 5.0 \times 10^{-14}$ J at a leakage–switching crossover of 400 Hz per synapse, a power-limited chip-wide throughput of 4.0×10^{12} events/s at a 100 mW budget, and a $7.4\times$ efficiency gain available from 90% adiabatic charging. The Landauer floor (2.87×10^{-21} J at 300 K) and the quantum speed limit are numerically inert at all realistic operating points—the fundamental energy–speed trade-off is flat from DC to the RC limit—so observed trade-offs in real hardware are engineering effects. The framework is falsifiable through measurable scaling exponents and a predicted crossover frequency testable on existing hardware within twelve months.

1. Introduction

The central promise of neuromorphic computing—brain-like efficiency in silicon—remains largely an empirical claim. Individual chips demonstrate impressive energy-per-spike figures, but there is no accepted theory that predicts, from physical first principles, how energy efficiency must degrade as computational speed or network size increases. This gap matters practically: chip architects tune operating points by trial and error, and claims of "brain-scale efficiency" are rarely checked against hard thermodynamic bounds.

This paper constructs such a theory. Our approach is quantum-inspired in a specific and defensible sense: we do not claim quantum computation occurs in neuromorphic hardware, but we import the variational and bound-structure of quantum thermodynamics—speed limits, entropy-production floors, and geometric dissipation functionals—into a classical non-equilibrium description. This strategy mirrors the successful use of quantum-inspired classical algorithms and simulators, where quantum formalism yields practical classical benefits without quantum hardware [1,7].

The core modeling move is to treat the neuromorphic processor as a non-equilibrium thermodynamic system in a driven steady state. Spiking activity is a probability flux through configuration space, sustained against dissipation by continuous power input, exactly analogous to the non-equilibrium steady states analyzed for thermally driven micromachines [8]. Dissipation is encoded in a Rayleigh-type potential relating thermodynamic forces and fluxes [2], and the strongly nonlinear, far-from-equilibrium character of spiking dynamics is handled with a local reduction to compact dynamical form in the spirit of Nambu non-equilibrium thermodynamics [3].

Our contributions are: (i) a three-term dissipation model for a synaptic event (Landauer erasure, non-adiabatic switching, leakage); (ii) an explicit derivation of the energy–speed frontier with all arithmetic shown; (iii) numerical evaluation for a representative 10^6 -synapse processor; and (iv) falsifiable scaling predictions testable within twelve months. Section 2 surveys the literature. Section 3 presents the model. Section 4 carries out the derivations. Section 5 reports results. Section 6 discusses limitations, falsifiability, and an explicit divergence resolution concerning error-correction overheads. Section 7 concludes.

2. Background and Related Work

Our model sits at the intersection of three literatures: quantum-inspired classical computing, geometric non-equilibrium thermodynamics, and the fundamental physical limits of computation.

Quantum-inspired classical computation. Tang examined quantum-inspired algorithms for recommendation systems and low-rank linear systems, showing that although these classical methods achieve exponential asymptotic speedups over prior classical algorithms, they carry hefty polynomial overheads relative to true quantum algorithms [1]. The lesson we adopt is methodological: formal structures borrowed from quantum theory must be checked against concrete resource arithmetic—precisely the discipline we apply to thermodynamic bounds. Complementing this, classical circuit networks have been shown to simulate Schrödinger dynamics and topological physics through the circuit-Laplacian correspondence, implementing genuine quantum-inspired information processing without quantum hardware [7]. Our model is the thermodynamic analogue: quantum-derived bounds applied to classical dissipative hardware.

Geometric non-equilibrium thermodynamics. Modern frameworks compare dissipation mechanisms—Rayleigh dissipation potentials, dissipative d'Alembert formulations, and gradient dynamics—within a unified geometric setting for multiscale systems [2]. We use the Rayleigh potential as our constitutive dissipation law because it yields a quadratic force–flux relation that is analytically tractable and empirically supported for resistive and capacitive electronic elements. For far-from-equilibrium, strongly nonlinear regimes, Nambu non-equilibrium thermodynamics provides local reduction of complex dynamics to a compact bracket form [3]; we invoke this to justify treating spiking dynamics as locally reduced two-variable dynamics (membrane voltage, synapse efficacy) even though the full system is high-dimensional.

Non-equilibrium steady states, damage, and criticality. The statistical mechanics of non-equilibrium damage phenomena develops a Gibbs-like formalism for non-equilibrium states and applies it to fiber-bundle models with thermal noise and fiber decay [4]. This is directly relevant to neuromorphic reliability: synapse degradation under sustained spiking is formally analogous to fiber failure under load, and we borrow the two-regime structure (fluctuation-dominated versus decay-dominated) to model device wear. Extended irreversible thermodynamics supplies the framework for critical behavior in non-equilibrium systems, where transport coefficients diverge near instability [6]; we use it to flag the breakdown of our linear force–flux assumption near synchronization transitions.

Coupled transport processes. A non-equilibrium thermodynamics model for combined adsorption and diffusion in micro- and nanopores shows that two coupled processes (diffusion-type and Langmuir-type dynamics) can be treated as diffusion in an effective landscape [5]. Our model of charge transport coupled to synaptic state update is structurally identical: ionic/electronic diffusion coupled to a saturating state variable, and we exploit that analogy to justify the coupled-equation structure of Section 3.

Micromachine steady states. The thermally driven three-sphere micromachine admits an exact non-equilibrium steady-state distribution with nonzero probability flux in configuration space [8]. This provides the cleanest available template for our central object: the spike-trajectory flux through the processor's phase space, whose circulation measures computational work done per unit time.

Fundamental limits. The physics-of-computation literature establishes the canonical bounds: the Landauer bound on erasure, the Margolus–Levitin theorem on evolution rate per unit energy, the Bremermann and Bekenstein limits, and—critically—the observation that error-correction overheads of 10^2 – 10^3 physical operations per logical operation multiply thermodynamic cost by the same factor in fault-tolerant quantum computing [12,10,11]. Thermodynamic viability must be assessed at the system level, not the device level [9]. Our Pareto-frontier analysis is the neuromorphic instantiation of that system-level accounting; notably, neuromorphic hardware avoids the quantum error-correction multiplier entirely, a quantitative advantage we examine explicitly (and whose treatment differs across source analyses; see Appendix A).

3. Methods

3.1 System model

We model a neuromorphic processor as N_s *synapses* and N_c *neurons*, each synaptic event (spike delivered to a synapse, weight applied, possible output spike) being the elementary computational operation. State variables per neuron: membrane voltage v and recovery variable u ; per synapse: efficacy w . Dynamics follow leaky integrate-and-fire form, treated as the locally reduced dynamics justified by the Nambu-reduction argument [3]:

$$dv/dt = (-v + I_{syn})/\tau_m, dw/dt = -w/\tau_w + \delta(t_{spike}) \Delta w,$$

with τ_m the *membrane time constant* and τ_w the *synaptic retention time constant*. This coupled two-timescale structure mirrors the diffusion–adsorption coupling of [5]: a transport process (charge integration) coupled to a saturating state variable (synaptic efficacy). In steady state the device carries a nonzero probability flux in configuration space, in direct analogy with the micromachine of [8]; this flux is the thermodynamic signature of ongoing computation.

3.2 Dissipation decomposition

Energy per synaptic event E_{ev} is decomposed into three terms:

$$E_{ev} = E_{Landauer} + E_{switch} + E_{leak}.$$

(a) Landauer term. Each synaptic event logically erases at least the information in the arriving spike's arrival-time uncertainty. We take the minimal erasure of one bit at temperature T : $E_{Landauer} = k_B T \ln 2$.

(b) Switching term. Charging and discharging the synaptic capacitance C_s through voltage swing ΔV dissipates, for non-adiabatic switching, $E_{switch} = C_s \Delta V^2$ (full CV^2 dissipation; adiabatic charging reduces this by factor $(1 - \alpha_{ad})$, the adiabaticity fraction, treated as a parameter). This is the Rayleigh-dissipation term: the Rayleigh potential $R = (1/2) g \dot{q}^2$ with conductance g yields quadratic dissipation in flux, consistent with [2].

(c) Leakage term. Static leakage power P_{leak} per synapse integrated over the event period gives $E_{leak} = P_{leak} / f_{ev}$, where f_{ev} is the per-synapse event rate.

3.3 Speed limit

The Margolus–Levitin theorem bounds the orthogonalization time of a system with mean energy E above ground: $t \geq \hbar/(2E)$ [12]. For a classical switching event driven by energy E_{switch} , we adopt the classical analogue: the minimum transition time through an RC network is

$$t_{sw} = R_{on} C_s \ln(V_{th}/(V_{th} - \Delta V)),$$

and the corresponding maximum event rate is $f_{max} = 1/t_{sw}$. This is the quantum-speed-limit structure (energy–time trade-off) in its classical RC limit, the same classical–quantum correspondence exploited in circuit-based quantum simulators [7].

3.4 Steady-state flux and the Pareto frontier

In the non-equilibrium steady state, total power $P_{tot} = N_s f_{ev} E_{ev}$ sustains a probability flux J through phase space [8]. Define the dimensionless dissipation ratio

$$\rho = E_{switch} / (k_B T \ln 2),$$

which measures how far above the Landauer floor the hardware operates. The frontier in the (f_{ev}, E_{ev}) plane follows from the RC speed limit: *E_{switch} is speed-independent for $f_{ev} \leq f_{max}$, and no solution exists for $f_{ev} > f_{max}$. The interesting structure is therefore in the (f_{ev}, ρ) plane and in the system-level constraint $P_{tot} \leq P_{budget}$.*

3.5 Numerical parameters

All input numbers, with sources:

- $T = 300$ K (standard room-temperature assumption).
- $k_B = 1.380649 \times 10^{-23}$ J/K; $\hbar = 1.054571817 \times 10^{-34}$ J·s (SI defined constants).
- $C_s = 100$ fF = 1.0×10^{-13} F (stated assumption: representative analog-CMOS synapse capacitance).
- $\Delta V = 0.5$ V (stated assumption: subthreshold analog swing).
- $R_{on} = 10$ k Ω (stated assumption: access-transistor on-resistance).
- $P_{leak} = 10$ pW per synapse (stated assumption: subthreshold leakage).
- $N_s = 10^6$ synapses (stated assumption: representative mid-scale chip).
- $f_{ev} = 1$ Hz per synapse baseline, swept to 10^4 Hz (stated assumption).
- $P_{budget} = 100$ mW (stated assumption: embedded-power envelope).

4. Analysis

4.1 Landauer term

$E_{Landauer} = k_B T \ln 2 = (1.380649 \times 10^{-23})(300)(0.693147)$. Step 1: $k_B T = 4.141947 \times 10^{-21}$ J. Step 2: $\times \ln 2 \rightarrow E_{Landauer} = 2.87 \times 10^{-21}$ J (0.693 $k_B T$).

4.2 Switching term

$E_{switch} = C_s \Delta V^2 = (1.0 \times 10^{-13})(0.25) = 2.5 \times 10^{-14}$ J. $\rho = 2.5 \times 10^{-14} / 2.87 \times 10^{-21} = 8.7 \times 10^6$. The representative synapse operates ~ 8.7 million times above the Landauer floor ($8.7 \times 10^6 k_B T \ln 2$). With $\alpha_{ad} = 0.9$: $E_{switch}^{eff} = 2.5 \times 10^{-15}$ J, $\rho = 8.7 \times 10^5$.

4.3 Switching time and speed limit

With $V_{th} = 0.9 \Delta V = 0.45$ V: $\ln(0.45/0.05) = \ln 9 = 2.19722$. $t_{sw} = (10^4)(1.0 \times 10^{-13})(2.19722) = 2.20 \times 10^{-9}$ s ≈ 2.2 ns; $f_{max} = 4.6 \times 10^8$ events/s. Cross-check against the Margolus–Levitin analogue: $t \geq \hbar/(2E_{switch}) = 1.0546 \times 10^{-34}/(5.0 \times 10^{-14}) = 2.1 \times 10^{-21}$ s—negligible compared to the RC limit. Neuromorphic speed is RC-limited, not quantum-limited, by $\sim 10^{12}$; the quantum speed-limit formalism is structurally useful but numerically inert here, an honest negative finding.

4.4 Leakage term and crossover

$E_{leak} = P_{leak}/f_{ev}$: 1.0×10^{-11} J at 1 Hz; 1.0×10^{-13} J at 10^2 Hz; 1.0×10^{-15} J at 10^4 Hz. The switching and leakage terms cross when $P_{leak}/f_{ev} = C_s \Delta V^2$: $f_{cross} = 10^{-11}/2.5 \times 10^{-14} = 400$ Hz. Below 400 Hz per synapse, leakage dominates; above, switching dominates. At $f_{ev} = 400$ Hz: $E_{ev} \approx 5.0 \times 10^{-14}$ J (the Landauer term is negligible at every operating point, by $\rho \approx 8.7 \times 10^6$).

4.5 System-level power and the frontier

$P_{tot} = N_s f_{ev} E_{ev}$. In the switching-dominated regime: $P_{tot} = 10^6 \times (2.5 \times 10^{-14} f_{ev} + 10^{-11})$. At $f_{ev} = 1$ Hz: ≈ 10 μ W; at 10^2 Hz: 12.5 μ W; at 10^3 Hz: 35 μ W; at 10^4 Hz: 260 μ W. Power-budget check: setting $P_{tot} = 0.1$ W gives $2.5 \times 10^{-14} f_{ev} = 10^{-7} - 10^{-11} \approx 9.999 \times 10^{-8}$, so $f_{ev} = 4.0 \times 10^6$ Hz per synapse—i.e., 4.0×10^{12} synaptic events/s chip-wide—comfortably below $f_{max} = 4.6 \times 10^8$ Hz. The binding constraint at this scale is power, not switching physics.

4.6 Adiabatic improvement

With $\alpha_{ad} = 0.9$, the crossover moves to $f_{cross} = 10^{-11}/2.5 \times 10^{-15} = 4000$ Hz, and chip-wide power at 10^4 Hz becomes $10^6 \times (2.5 \times 10^{-11} + 10^{-11}) = 35$ μ W—a 7.4 $\times$ reduction ($260/35 = 7.43$).

4.7 Redundancy overhead analogy

Following the system-level lesson of fault-tolerant quantum computing, where 10^2 – 10^3 physical operations per logical operation multiply thermodynamic cost [12,10], suppose neuromorphic robustness requires redundancy factor $r = 100$ (stated assumption: triple-modular-style voting with spare synapses). Chip-wide logical throughput at fixed power falls by r : at 100 mW, logical rate = 4.0×10^{10} events/s, and energy per logical event rises to 2.5×10^{-12} J = 8.7×10^8 $k_B T \ln 2$. Results scale linearly in r ; for $r \in [10, 1000]$, logical energy $\in [2.5 \times 10^{-13}, 2.5 \times 10^{-11}]$ J.

5. Results

All numbers are computed in Section 4 from stated inputs; none are empirical measurements.

1. **Landauer floor:** 2.87×10^{-21} J (0.693 $k_B T$) per synaptic event at 300 K.
2. **Representative switching energy:** 2.5×10^{-14} J (8.7×10^6 $k_B T \ln 2$), $\rho = 8.7 \times 10^6$; with 90% adiabatic recovery, 2.5×10^{-15} J ($\rho = 8.7 \times 10^5$).
3. **RC speed limit:** $t_{sw} = 2.2$ ns, $f_{max} = 4.6 \times 10^8$ events/s per synapse; the Margolus–Levitin analogue gives 2.1×10^{-21} s, confirming the RC limit binds by $\sim 10^{12}$.
4. **Leakage–switching crossover:** $f_{cross} = 400$ Hz per synapse (non-adiabatic); 4000 Hz with $\alpha_{ad} = 0.9$.
5. **Minimum energy per event:** $\approx 5.0 \times 10^{-14}$ J $\approx 1.74 \times 10^7$ $k_B T \ln 2$ at the crossover (non-adiabatic)—the model's predicted optimal operating point.
6. **Chip-level power curve ($N_s = 10^9$):** 10 μ W (1 Hz/synapse), 12.5 μ W (10^2 Hz), 35 μ W (10^3 Hz), 260 μ W (10^4 Hz); with $\alpha_{ad} = 0.9$, 35 μ W at 10^4 Hz (7.4 $\times$ improvement).
7. **Power-budget-limited throughput:** at 100 mW, 4.0×10^6 events/s per synapse (4.0×10^{12} chip-wide), two orders of magnitude below the RC limit; power, not device physics, binds.
8. **Projection (labeled):** with redundancy $r = 100$, energy per logical event rises to 2.5×10^{-12} J and logical throughput falls to 4.0×10^{10} events/s at 100 mW (range 2.5×10^{-13} – 2.5×10^{-11} J for $r \in [10, 1000]$).
9. **Flat fundamental trade-off:** because E_{switch} is speed-independent below f_{max} , the fundamental energy–speed curve is flat from DC to $\sim 4.6 \times 10^8$ Hz; observed speed–efficiency slopes in neuromorphic hardware are engineering artifacts (charging, leakage, routing), not fundamental physics.

6. Discussion

Limitations. The model's weakest link is the parameter set. $C_s = 100 \text{ fF}$, $R_{on} = 10 \text{ k}\Omega$, $\Delta V = 0.5 \text{ V}$, and $P_{leak} = 10 \text{ pW}$ are stated assumptions representative of analog CMOS, not measurements of any specific chip; memristive or spintronic synapses would shift E_{switch} by orders of magnitude in either direction. The Rayleigh quadratic force-flux law [2] breaks down near criticality—extended irreversible thermodynamics predicts divergent transport coefficients near non-equilibrium instabilities [6]—so our frontier is invalid near synchronization transitions. The Nambu-style local reduction [3] is justified only away from strong-coupling regimes; correlated firing is outside the model. The single-bit Landauer erasure assumption is conservative in one direction (analog spike timing may erase less) and optimistic in another (reset, routing, and conversion are ignored).

Failure modes and falsifiability. The central falsifiable prediction is the leakage-switching crossover at $f_{cross} = 400 \text{ Hz}$ (non-adiabatic parameters): a chip measured to have rate-independent energy per event well below 400 Hz, or a crossover displaced by more than an order of magnitude from the value implied by its measured C_s , ΔV , and P_{leak} , falsifies the three-term decomposition as applied. If measured energy per spike falls below our E_{switch} floor, the capacitance assumption is wrong, not the framework. If energy per spike rises with firing rate beyond the RC-predicted step, additional dissipation channels must be added as further Rayleigh terms. The redundancy projection is falsified if system-level energy per logical operation shows no overhead factor relative to device-level energy per event.

Arguing against ourselves. A skeptic could say the model is elaborate bookkeeping around the trivial fact that CV^2 dominates: the Landauer term, the quantum speed limit, and the geometric formalism contribute nothing numerically ($\rho \approx 8.7 \times 10^6$; quantum bound inert by 10^{12}). This is partly fair—and it is itself a result, echoing the cautionary lesson of quantum-inspired algorithms whose formal elegance outran practical advantage [1]. Our defense is that the framework identifies *which* bound binds and *where the frontier bends* (the 400 Hz crossover, the power-limited regime at $4 \times 10^6 \text{ Hz/synapse}$), which bare CV^2 arithmetic does not. The damage-physics analogy [4] suggests a further open direction: synapse wear under sustained flux may impose a lifetime-rate trade-off we have not modeled. Open questions include: the correct dissipation potential for subthreshold adiabatic charging; whether critical slowing near synchronization alters the frontier's slope [6]; and whether coupled transport-state dynamics [5] predict additional cross-terms measurable as rate-dependent leakage.

On error-correction overheads. One source analysis applied a combined quantum-inspired algorithmic and error-correction multiplier ($\eta \approx 3.16 \times 10^4$) directly to neuromorphic operations, yielding a 0.5 pJ optimum under a 0.5 W / 1 TOPS scenario. We adopt the opposite convention—neuromorphic hardware carries no QEC multiplier, only an optional classical redundancy factor r —because B and C converge on the structural argument that QEC overheads are specific to fault-tolerant quantum architectures [10,11,12]; the conflict is documented in Appendix A rather than silently resolved.

7. Conclusion

We constructed a non-equilibrium thermodynamic model of neuromorphic computation that treats the processor as a driven dissipative system sustaining computational flux, decomposes energy per synaptic event into Landauer, switching, and leakage terms, and derives an explicit energy-speed frontier. For representative analog-CMOS parameters, the model predicts a minimum energy per synaptic event of $\sim 5.0 \times 10^{-14} \text{ J}$ at a 400 Hz crossover, a power-limited chip-wide throughput of $4.0 \times 10^{12} \text{ events/s}$ at 100 mW for 10^6 synapses, and a 7.4× efficiency gain available from 90% adiabatic charging. The Landauer floor and quantum speed limits, while structurally informative, are numerically inert at current operating points—efficiency gains must come from adiabaticity and leakage reduction, not from approaching fundamental bounds. The framework's crossover and scaling predictions

are testable on existing hardware within twelve months, and its system-level accounting, informed by the redundancy overheads known from fault-tolerant quantum computing, provides the honest boundary assessment that neuromorphic efficiency claims have lacked.

References

[1] Quantum-inspired algorithms in practice. arXiv:1905.10415v3. <https://arxiv.org/abs/1905.10415v3> [2] Comparison of some geometric frameworks for dissipative evolution in multiscale non-equilibrium thermodynamics. arXiv:2512.05168v2. <https://arxiv.org/abs/2512.05168v2> [3] Reduction of Complex Dynamics in Far-from-equilibrium Systems: Nambu Non-equilibrium Thermodynamics. arXiv:2508.19455v3. <https://arxiv.org/abs/2508.19455v3> [4] Non-equilibrium statistical mechanics of non-equilibrium damage phenomena. arXiv:0905.0292v3. <https://arxiv.org/abs/0905.0292v3> [5] A non-equilibrium thermodynamics model for combined adsorption and diffusion processes in micro- and nanopores. arXiv:1204.3841v1. <https://arxiv.org/abs/1204.3841v1> [6] Non-equilibrium critical behavior : An extended irreversible thermodynamics approach. arXiv:0908.2161v1. <https://arxiv.org/abs/0908.2161v1> [7] Engineering topological states and quantum-inspired information processing using classical circuits. arXiv:2409.09919v2. <https://arxiv.org/abs/2409.09919v2> [8] Non-equilibrium probability flux of a thermally driven micromachine. arXiv:1905.06796v2. <https://arxiv.org/abs/1905.06796v2> [9] DOI 10.5281/zenodo.18036068. QNFO: Thermodynamic Viability and the Universality of Feynman Matter. [10] DOI 10.5281/zenodo.17955898. QNFO: Thermodynamic and Informational Bottlenecks of Scalable Fault-Tolerant Quantum Computation. [11] DOI 10.5281/zenodo.17937531. QNFO: Thermodynamic and Quantum Constraints on Scalable Quantum Computing. [12] DOI 10.5281/zenodo.22753039. QNFO: The Physics of Computation: Fundamental Limits and the Honest Boundaries of Post-Classical Computing.

Appendix A. Divergence report

D1. Applicability of quantum error-correction overhead to neuromorphic hardware. Draft A applies a combined multiplier $\eta = 10^2 \times \eta_{\text{QEC}} \approx 3.16 \times 10^4$ (algorithmic prefactor from [1] times median QEC overhead from [10]) directly to neuromorphic logical operations, obtaining an optimal E_{op} of 5.0×10^{-13} J under a 0.5 W / 1 TOPS scenario. Drafts B and C apply no QEC multiplier to neuromorphic operations, arguing (B, §2; C, §2/§4.8) that the 10^2 – 10^3 overhead of [10,11,12] is specific to fault-tolerant quantum architectures, and that neuromorphic hardware avoids it entirely; C instead introduces an optional classical redundancy factor r as a labeled assumption. *Underlying convention disagreement:* whether "quantum-inspired" algorithms running on classical neuromorphic hardware inherit quantum error-correction costs. *Resolution:* the main text adopts the B/C convention (no QEC multiplier; optional classical redundancy r), because two independent drafts converge on the structural argument and A's application of a quantum-QEC factor to non-quantum hardware lacks a mechanism. A's 0.5 pJ / 1 TOPS scenario is retained as a documented alternative convention, not as the main result.

D2. Power budget and scale. Draft A assumes $P_{\text{max}} = 0.5$ W and $v_{\text{target}} = 10^{12}$ ops/s (chip-level TOPS framing). Draft C assumes $P_{\text{budget}} = 100$ mW with $N_{\text{s}} = 10^6$ synapses and per-synapse event rates. Draft B uses per-channel and per-chip floor framing with $N = 10^6$ synapses at 1% activity. *Underlying disagreement:* chip class (high-performance module vs. embedded envelope) and accounting unit (logical TOPS vs. synaptic events). *Resolution:* the main text adopts C's embedded-envelope framing (100 mW, 10^6 synapses) because it yields a fully explicit, self-consistent parameter set; A's scenario is reported here for completeness. Under A's convention, the power-budget constraint gives $E_{\text{op}} \leq P/v_{\text{target}} = 0.5/10^{12} = 5.0 \times 10^{-13}$ J, with the Margolus–Levitin ceiling $v_{\text{max}} = 2P/\hbar = 1.51 \times 10^{33}$ ops/s non-binding—consistent with the main text's finding that quantum limits are inert.

D3. Optimal operating point. Draft A reports $E_{\text{op}} = 5.0 \times 10^{-13} \text{ J}$ (0.5 pJ) as the optimum of a weighted cost functional. Draft C reports $\approx 5.0 \times 10^{-14} \text{ J}$ at the 400 Hz crossover. Draft B reports a practical floor of $\sim 110 \text{ fJ/op}$ (range $10 \text{ fJ} - 1 \text{ pJ}$) and a fundamental floor of $2.87 \times 10^{-21} \text{ J}$. Underlying disagreement: different objective functions (weighted energy/speed cost vs. minimum of a three-term decomposition vs. leakage+I/O practical floor) and different parameter sets. Resolution: * the main text adopts C's three-term decomposition optimum, which follows from the explicitly stated parameter set; the spread across drafts ($5 \times 10^{-14} - 5 \times 10^{-13} \text{ J}$) reflects parameter assumptions, not physical disagreement, and all drafts agree the value lies $\geq 10^7$ above the Landauer floor.

D4. Role of the Margolus–Levitin bound. All three drafts agree the quantum speed limit does not bind (A: ceiling $1.51 \times 10^{33} \text{ ops/s}$ far above target; B: crossover at 58 fs; C: RC limit binds by $\sim 10^{12}$). CONVERGENT; no conflict.

Appendix B. Claim attribution

Claim	Drafts	Status
C1. Landauer floor = $k_B T \ln 2 \approx 2.87 \times 10^{-21} \text{ J}$ at 300 K	A, B, C	CONVERGENT
C2. Margolus–Levitin / quantum speed limit is numerically inert at realistic neuromorphic operating points	A, B, C	CONVERGENT
C3. Rayleigh dissipation potential is the appropriate constitutive dissipation law	A, B, C	CONVERGENT
C4. Nambu non-equilibrium thermodynamics justifies local reduction of spiking dynamics	A, B, C	CONVERGENT
C5. Quantum-inspired algorithms carry polynomial overheads eroding practical advantage [1]	A, B, C	CONVERGENT
C6. Classical circuits can host quantum-inspired information processing [7]	A, B, C	CONVERGENT
C7. QEC overhead of $10^2 - 10^3$ multiplies thermodynamic cost in fault-tolerant quantum computing [10,11,12]	A, B, C	CONVERGENT
C8. QEC/algorithmic overhead multiplier applies to neuromorphic logical operations ($\eta \approx 3.16 \times 10^4$)	A	SINGLE (contra B, C) → DIVERGENT (D1)
C9. Optimal $E_{\text{op}} = 5.0 \times 10^{-13} \text{ J}$ under 0.5 W / 1 TOPS scenario	A	SINGLE (contra B, C) → DIVERGENT (D2, D3)
C10. Power budget 0.5 W, target 10^{12} ops/s	A	SINGLE → DIVERGENT (D2)
C11. Power budget 100 mW, $N_s = 10^6$ synapses, per-synapse event-rate framing	C (B compatible)	CONVERGENT (B, C)
C12. Three-term decomposition (Landauer + switching + leakage) with crossover $f_{\text{cross}} = 400 \text{ Hz}$	C (B qualitatively: leakage/I/O dominate)	CONVERGENT (B, C)
C13. Fundamental energy–speed trade-off is flat across realistic clock rates; observed trade-offs are engineering effects	B, C	CONVERGENT
C14. Practical floor set by leakage and I/O, not fundamental bounds; $\sim 10^4 - 10^{10}$ headroom above Landauer	B, C	CONVERGENT
C15. 12-month falsifiable predictions on existing hardware	B, C	CONVERGENT
C16. Adiabatic charging yields $\sim 7.4\times$ efficiency gain at 90% recovery	C	SINGLE
C17. Classical redundancy factor r multiplies logical energy per event (system-level lesson of QEC)	C (A structurally similar via η)	CONVERGENT (A, C, different magnitudes; see D1)
C18. Damage/fiber-bundle formalism as template for synapse wear	A, B, C	CONVERGENT
C19. Coupled adsorption–diffusion analogy for synapse state dynamics	A, B, C	CONVERGENT

Claim	Drafts	Status
C20. Critical slowing near non-equilibrium instabilities breaks linear force–flux assumptions	A, B, C	CONVERGENT