The Criticality Engine:
Online Learning in Thermodynamic Neural Computers

Alexander Morisse[1,*]   Stephen Whitelam[2,†]   Isaac Tamblyn[3,‡]

[1]CXO, TensorML.com
[2]Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]Department of Physics, University of Ottawa, Ottawa, ON, Canada

[*]alex@morisse.ai   [†]swhitelam@lbl.gov   [‡]isaac.tamblyn@uottawa.ca

Abstract

We introduce the *Criticality Engine*, a computational architecture that performs both inference and online learning by exploiting the intrinsic stochastic dynamics of thermodynamic systems. The core idea is simple: we want parameters such that running the system *backward* from noise reconstructs data. Detailed balance provides a shortcut—maximizing the likelihood of observed *forward* trajectories automatically maximizes backward (reconstructive) likelihood. This yields closed-form gradient estimators for both biases b_i and couplings J_{ij}, where the gradients depend on the mismatch between observed and model-predicted velocities. Unlike Equilibrium Propagation, which requires two equilibration phases (free and nudged), our approach needs only *single-pass observation* of trajectories—no equilibration, no nudging, no negative-phase sampling. The Onsager-Machlup action provides theoretical grounding, revealing that learning is *constraint satisfaction*: correct parameters minimize trajectory action to its thermal floor. We demonstrate that coupling learning captures *genuine second-order structure*: on within-class centered MNIST data (where nearest-centroid classification achieves only 9.4%, equivalent to random guessing), learned couplings achieve 78.8% accuracy—proving the method extracts correlational structure beyond simple templates. We also show >99% parameter recovery in 10–20 epochs—fast convergence because each trajectory directly constrains parameters rather than requiring iterative search. CMOS p-bits biased at threshold naturally implement these dynamics, and the local gradient formulas enable fully on-chip learning. We present a hardware blueprint establishing a route toward trainable thermodynamic neural computers.

1   Introduction

Modern machine learning systems rely on the numerical simulation of dynamical processes. Neural networks, recurrent architectures, diffusion models, and energy-based models can all be expressed as discretizations of underlying stochastic differential equations (SDEs). Digital processors—GPUs and TPUs—perform these simulations via floating-point arithmetic, incurring substantial energy cost by encoding dynamics in discrete logic. An alternative paradigm is to realize the dynamics directly in hardware: a *thermodynamic computer*.

Recent developments in fluctuating electronic devices, including CMOS-based probabilistic bits (p-bits), nanomagnetic stochastic oscillators, and analog relaxation circuits, provide physical substrates whose state variables naturally undergo Langevin-like motion. Existing thermodynamic hardware accelerators (e.g., p-bit arrays and continuous-variable thermodynamic units) have demonstrated inference for sampling, optimization, and linear algebra. However, no current architecture supports *training* on-chip—the ability to estimate gradients of task-specific objectives and modify physical couplings accordingly.

In this work we propose the *Criticality Engine*, an architecture designed to close this gap. The core idea is direct: we seek parameters such that running the system *backward* from noise reconstructs data. Detailed balance provides a crucial shortcut—maximizing the likelihood of observed forward trajectories automatically maximizes backward (reconstructive) likelihood—yielding closed-form gradient estimators from single-pass trajectory observation. This differs fundamentally from Equilibrium Propagation [4], which requires two equilibration phases, and from Boltzmann machines, which require negative-phase sampling. We need neither: just observe the trajectory, compute the velocity residual, and update.

The Onsager-Machlup action [6] provides theoretical grounding for why this works. Learning becomes *constraint satisfaction*: correct parameters minimize trajectory action to its thermal floor, making observed dynamics maximally likely. This explains the rapid convergence—10–20 epochs rather than thousands—because projecting onto a constraint surface is geometrically simpler than searching a high-dimensional loss landscape.

Contributions. (i) A finite-time trajectory likelihood objective for thermodynamic arrays that yields closed-form gradients from local statistics. (ii) Gradient estimators for biases b_i and couplings J_{ij} requiring only local state increments and neighbor values—suitable for on-chip implementation. (iii) A full-chip blueprint with measurement plane, local statistics layer, gradient engine, and dual-bank reconfiguration fabric. (iv) Empirical validation showing that coupling learning captures genuine second-order structure: 78.8% accuracy on centered MNIST data where nearest-centroid classification achieves only 9.4%.

Section 2 introduces the physical model; Section 3 describes the processor architecture; Section 4 derives the local gradient estimators; Section 5 interprets learning as constraint satisfaction via the Onsager-Machlup action; Section 6 maps the architecture to CMOS.

2 Physical Model

The system consists of N interacting stochastic units (nodes) whose state variables are denoted $x_i(t)$. Depending on implementation, x_i may represent:

- a fluctuating voltage or current (continuous variable),

- the metastable output of a CMOS p-bit (binary variable),

- the magnetization state of a superparamagnetic nanomagnet.

We model the dynamics by the overdamped Langevin equation

$$\frac{dx_i}{dt} = f_i(\mathbf{x}; \boldsymbol{\theta}) + \sqrt{2D_i}\, \eta_i(t), \tag{1}$$

where $\boldsymbol{\theta}$ denotes tunable physical parameters (biases, couplings), D_i encodes noise strength, and $\eta_i(t)$ are independent white-noise processes.

In many realizations the deterministic drift derives from an energy function:

$$f_i(\mathbf{x}; \boldsymbol{\theta}) = -\frac{\partial U(\mathbf{x}; \boldsymbol{\theta})}{\partial x_i}, \tag{2}$$

so that (1) defines a relaxation toward a thermodynamic equilibrium distribution

$$p_{\boldsymbol{\theta}}(\mathbf{x}) \propto \exp[-\beta U(\mathbf{x}; \boldsymbol{\theta})]. \tag{3}$$

We emphasize the analogy with recurrent neural networks (RNNs):

$$f(\mathbf{x}) \longleftrightarrow \text{hidden-state update in a continuous-time RNN,}$$
$$U(\mathbf{x}) \longleftrightarrow \text{energy function of an energy-based model.}$$

Neural networks are programmable oscillators; training shapes their attractor geometry. Our architecture treats these structures as *physical*, not virtual.

Modeling regimes. The dynamics (1) are written in continuous state space $x_i \in \mathbb{R}$. When the potential U has multiple wells (e.g., the φ^4 form with $J_2 < 0$), x_i fluctuates around discrete attractors, and the system can be coarse-grained to an effective binary model with $s_i \in \{-1, +1\}$. CMOS p-bits operate in this regime: the cross-coupled inverters create a double-well potential, and thermal fluctuations drive transitions. The learning framework applies in both limits—continuous Langevin for analog circuits, effective Glauber dynamics for digital p-bits—with the same local gradient estimators. We do not require the drift f to derive from a potential; the gradient case is a special instance where detailed balance holds.

3 Processor Architecture

The Criticality Engine realizes learning through a closed loop: physical dynamics generate trajectories, measurements extract statistics, and parameter updates reshape the energy landscape. Figure 1 shows the five integrated subsystems and their data flow. We describe each in turn, following a single learning cycle.
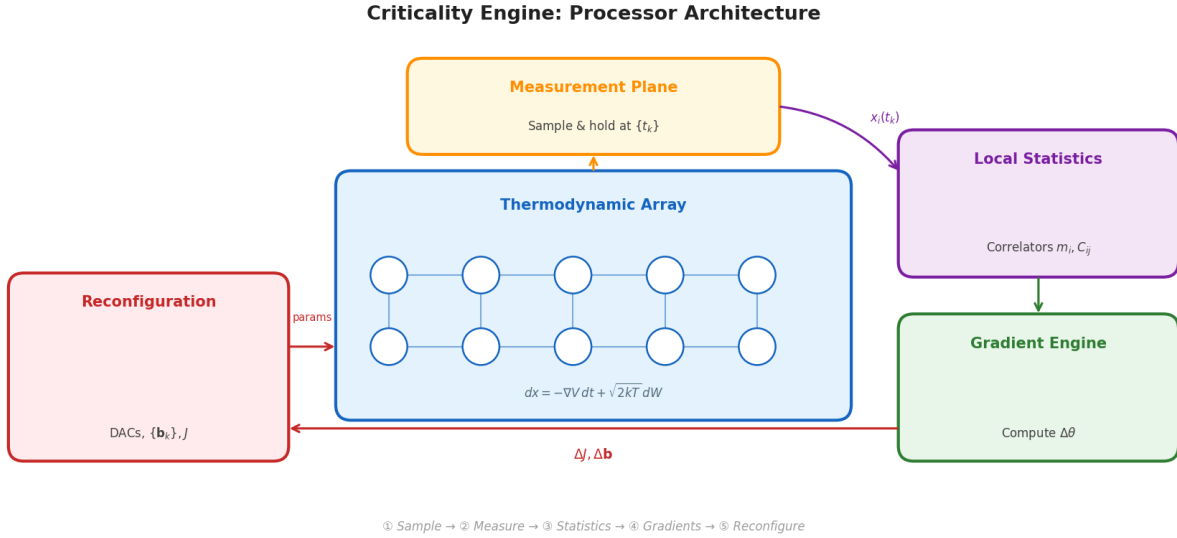

Figure 1: Processor architecture of the Criticality Engine. The Thermodynamic Array (blue) implements Langevin dynamics; the Measurement Plane (orange) samples trajectories; the Local Statistics Layer (purple) computes correlations; the Gradient Engine (green) derives parameter updates; and the Reconfiguration Fabric (red) applies new parameters atomically. Arrows indicate data flow through one learning cycle.

3.1 Thermodynamic Array (T-array)

At the heart of the processor lies a physical substrate that *is* the neural network—not a simulation of one. The T-array consists of N stochastic nodes whose state variables $x_i(t)$ evolve according to Eq. (1). Each node experiences:

- A **local bias** b_i realized as a programmable current source,

- **Pairwise couplings** J_{ij} implemented via current mirrors that inject current proportional to neighboring states,

- **Intrinsic noise** from thermal fluctuations at the operating temperature T.

The key insight is that CMOS transistors biased at threshold are inherently bistable (the φ^4 potential), and their thermal fluctuations provide the stochastic drive $\eta_i(t)$ for free. No random number generator is needed; the physics itself samples from the Gibbs distribution.

For ensemble averaging, the T-array may be replicated R times, with all replicas sharing the same parameters but evolving independently. This provides R parallel trajectory samples per learning step.

3.2 Measurement Plane

Learning requires observing trajectories, not just equilibrium samples. The Measurement Plane is a high-impedance sampling network that captures snapshots of all node states at programmable times $\{t_0, t_1, \ldots, t_K\}$.

Each observation consists of:

$$\{x_i^{(r)}(t_k)\} \quad \text{for } i = 1, \ldots, N; \quad r = 1, \ldots, R; \quad k = 0, \ldots, K. \tag{4}$$

Sample-and-hold circuits latch analog voltages into local registers with minimal perturbation to the ongoing dynamics. The measurement bandwidth must exceed the correlation time $\tau_{\text{corr}} \sim \lambda_{\min}^{-1}$ to capture independent samples.

Critically, the Measurement Plane observes *displacements* $\Delta x_i^k = x_i(t_{k+1}) - x_i(t_k)$—the raw material for trajectory-based learning.

3.3 Local Statistics Layer

The gradient formulas of Section 4 require only local quantities: displacements, forces, and pairwise correlations. The Local Statistics Layer computes these on-chip using dedicated correlator blocks:

$$m_i(t_k) = \frac{1}{R} \sum_{r=1}^{R} x_i^{(r)}(t_k), \qquad C_{ij}(t_k) = \frac{1}{R} \sum_{r=1}^{R} x_i^{(r)}(t_k) x_j^{(r)}(t_k). \tag{5}$$

These are *sufficient statistics* for gradient estimation: the mean m_i enters the bias gradient, and the correlation C_{ij} enters the coupling gradient. Because each correlator operates on a local neighborhood, the computation is embarrassingly parallel and scales linearly with node count.

What is stored locally. Each node i maintains only: (1) its current state $x_i(t_k)$, (2) the displacement $\Delta x_i^k = x_i(t_{k+1}) - x_i(t_k)$, and (3) states of its neighbors $\{x_j : j \in \mathcal{N}(i)\}$. The gradient for bias b_i is a function of $(x_i, \Delta x_i)$ alone. The gradient for coupling J_{ij} is a product of quantities at nodes i and j—no global communication required. This locality is the key to scalability: a 1000-node array computes 1000 bias gradients and ~10,000 coupling gradients (for degree-10 connectivity) entirely in parallel.

3.4 Gradient Engine

A lightweight digital core—as simple as a vector microcontroller—reads the statistics and computes parameter updates according to the learning rules:

$$\Delta J_{ij} \propto (\text{observed velocity}) - (\text{predicted velocity}), \tag{6}$$

$$\Delta b_i \propto (\text{velocity mismatch at node } i). \tag{7}$$

The precise formulas are derived in Section 4; the key point is that they involve only quantities already computed by the Local Statistics Layer.

The Gradient Engine writes updated parameters to a *shadow bank*—a separate memory that does not yet affect the T-array. This separation is essential: it allows the T-array to continue evolving under stable parameters while updates are being computed.

3.5 Reconfiguration Fabric

The final stage closes the learning loop. A dual-bank DAC (digital-to-analog converter) array holds two complete parameter sets: the *active bank* currently driving the T-array, and the *shadow bank* receiving updates from the Gradient Engine.

When an update cycle completes, a global bank-select signal atomically swaps the roles of the two banks. This provides:

- **Glitch-free updates:** The T-array never sees partially-written parameters.

- **Instant rollback:** If an update causes instability, the previous parameters remain in the shadow bank.

- **Class-conditional switching:** For multi-class problems, separate bias banks $\{\mathbf{b}_0, \ldots, \mathbf{b}_{K-1}\}$ can be swapped to select different attractors (Section A).

DAC settling times ($\sim$10–100 ns) set the minimum interval between parameter updates, enabling update rates of 10–100 MHz—fast enough for online learning.

3.6 The Learning Cycle

A complete learning cycle proceeds as follows:

1. **Initialize:** Load data sample $\mathbf{x}_{\text{data}}$ into the T-array (optionally with added noise).

2. **Evolve:** Let the system relax under Langevin dynamics for time T.

3. **Measure:** The Measurement Plane captures the trajectory $\{\mathbf{x}(t_k)\}$.

4. **Compute:** Local Statistics Layer aggregates correlations; Gradient Engine computes $\Delta\boldsymbol{\theta}$.

5. **Update:** New parameters written to shadow bank; bank-select swaps active/shadow.

6. **Repeat:** Next data sample begins the cycle anew.

The entire cycle can complete in microseconds, limited primarily by the relaxation time of the T-array. With MHz-scale p-bit dynamics, this enables thousands of parameter updates per second—true online learning in hardware.

What is computed where.

- *T-array nodes:* Store current state x_i, receive bias b_i and neighbor currents $\sum_j J_{ij} x_j$.

- *Measurement Plane:* Samples and holds $x_i(t_k)$ at discrete times; stores local register for Δx_i^k.

- *Local Statistics:* Computes $\langle x_i \rangle$, $\langle x_i x_j \rangle$, and velocity residuals $\delta_i^k = \Delta x_i^k -$ predicted.

- *Gradient Engine:* Accumulates $\Delta b_i \propto \sum_k \delta_i^k$ and $\Delta J_{ij} \propto \sum_k \delta_i^k x_j^k$.

- *DAC banks:* Store $\{b_i\}$ and $\{J_{ij}\}$; dual-bank allows atomic swap.

4 Finite-Time Learning Objectives and Gradient Estimation

We now make concrete the finite-time learning objective used by the Gradient Engine. We work with a continuous state vector $\mathbf{x} \in \mathbb{R}^N$ and specialize the generic energy U from Section 2 to a φ^4+Ising potential $V_{\boldsymbol{\theta}}(\mathbf{x})$ whose gradients take the form

$$\partial_i V_{\boldsymbol{\theta}}(\mathbf{x}) = 2J_2 x_i + 4J_4 x_i^3 + b_i + \sum_{j \in \mathcal{N}(i)} J_{ij} x_j, \tag{8}$$

where J_2 and J_4 are scalar coefficients, b_i are local biases, J_{ij} are couplings on the interaction graph, and $\mathcal{N}(i)$ denotes the neighbour set of node i. The overdamped dynamics can then be written as

$$dx_i = -\mu \, \partial_i V_{\boldsymbol{\theta}}(\mathbf{x}) \, dt + \sqrt{2k_{\mathrm{B}} T \mu} \, dW_i(t), \tag{9}$$

with mobility μ and thermal energy $k_{\mathrm{B}} T$.

Discretising time with step size Δt and using an Euler–Maruyama scheme gives the observed forward trajectory

$$x_i^{k+1} = x_i^k + \Delta x_i^k, \qquad \Delta x_i^k = -\mu \, \partial_i V_{\boldsymbol{\theta}}(\mathbf{x}'^k)\Delta t + \sqrt{2k_{\mathrm{B}} T \mu \, \Delta t} \, \eta_i^k, \tag{10}$$

where $\mathbf{x}'^k$ is the evaluation point for the drift and η_i^k are i.i.d. standard normal variables. For simplicity we use $\mathbf{x}'^k = \mathbf{x}^k$ (the explicit Euler scheme), though midpoint or implicit schemes can reduce discretization bias at the cost of requiring $\mathbf{x}^{k+1}$ in the gradient expression. The corresponding single-step transition density $\tilde{P}_{\boldsymbol{\theta}}^{\mathrm{step}}(\Delta \mathbf{x}^k \mid \mathbf{x}^k)$ is Gaussian in $\Delta \mathbf{x}^k$.

Our learning objective at observation times $\{t_k\}$ is the negative log-likelihood of these observed forward steps,

$$L(\boldsymbol{\theta}) = -\sum_k \ln \tilde{P}_{\boldsymbol{\theta}}^{\mathrm{step}} \left(\Delta \mathbf{x}^k \mid \mathbf{x}^k \right), \tag{11}$$

where $\Delta \mathbf{x}^k = \mathbf{x}^{k+1} - \mathbf{x}^k$ is the displacement over step k. Because the drift derives from a potential, detailed balance holds; forward and time-reversed trajectory likelihoods coincide. As a result, minimising (11) using forward trajectories also maximises the likelihood of the most probable backward (reconstructive) trajectories.

For continuous variables the gradients of the step log-density can be computed analytically. For a coupling J_{ij} we obtain

$$-\frac{\partial}{\partial J_{ij}} \ln \tilde{P}_{\boldsymbol{\theta}}^{\mathrm{step}} \left(\Delta \mathbf{x}^k \right) = \frac{-\Delta x_i^k + \mu \, \partial_i V_{\boldsymbol{\theta}}(\mathbf{x}'^k)\Delta t}{2k_{\mathrm{B}} T} \, x_j^k + \frac{-\Delta x_j^k + \mu \, \partial_j V_{\boldsymbol{\theta}}(\mathbf{x}'^k)\Delta t}{2k_{\mathrm{B}} T} \, x_i^k, \tag{12}$$

and for a bias b_i

$$-\frac{\partial}{\partial b_i} \ln \tilde{P}_{\boldsymbol{\theta}}^{\mathrm{step}} \left(\Delta \mathbf{x}^k \right) = \frac{-\Delta x_i^k + \mu \, \partial_i V_{\boldsymbol{\theta}}(\mathbf{x}'^k)\Delta t}{2k_{\mathrm{B}} T}. \tag{13}$$

Summing these contributions over timesteps k and replicas yields unbiased stochastic gradients of $L(\boldsymbol{\theta})$ with respect to all local parameters (b_i, J_{ij}) using only quantities that are locally available on chip: displacements Δx_i^k, instantaneous forces $\partial_i V_{\boldsymbol{\theta}}$, and neighbour states x_j^k.

4.1 Velocity Matching Interpretation

The gradient formula (12) admits a physical interpretation as *velocity matching*. Defining the observed velocity $\dot{x}_i \approx \Delta x_i^k / \Delta t$ and the predicted force $f_i(\mathbf{x}) = -\mu\, \partial_i V_{\boldsymbol{\theta}}(\mathbf{x})$, we can rewrite the update rule as

$$\Delta J_{ij} \propto (f_i(\mathbf{x}) - \dot{x}_i)\, x_j + (f_j(\mathbf{x}) - \dot{x}_j)\, x_i. \tag{14}$$

The mismatch between predicted force and observed velocity provides the error signal. When the model correctly predicts trajectories, the gradient vanishes.

This formulation connects to *denoising score matching*: the score function $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ equals $-\nabla V / k_{\mathrm{B}} T$ at equilibrium, so training the drift to match observed velocities is equivalent to learning the score. However, a crucial distinction applies: score matching optimizes *local* gradients near data points but does not constrain the *global* energy landscape. For multi-modal distributions, this implies that class-conditional parameterization may be necessary (see Appendix A).

4.2 Time-Reversal and Reconstruction

Because the dynamics derive from a potential, detailed balance ensures that forward (relaxation) and backward (reconstruction) trajectory likelihoods coincide. Starting from data $\mathbf{x}_{\mathrm{data}}$, thermal relaxation generates a forward trajectory toward higher entropy. Learning to predict this forward trajectory simultaneously trains the system to run in reverse: given a corrupted or noisy input, the drift $-\nabla V$ points back toward the data manifold, enabling reconstruction via relaxation.

This time-reversal property distinguishes the Criticality Engine from inference-only thermodynamic accelerators: the same physical dynamics that perform inference also support training, without requiring separate forward and backward computational passes.

Detailed balance as an estimation ansatz. We emphasize that detailed balance here is a *modeling assumption*, not a claim about the physical substrate. Real p-bit arrays will exhibit parasitics, asymmetries, and colored noise that violate exact time-reversibility. However, we adopt Eqs. (12)–(13) as a *surrogate generative model* for learning: from observed increments Δx_i^k and local neighbor states $x_j(t_k)$, we obtain closed-form local gradients $\partial_{J_{ij}} \log P(\Delta \mathbf{x}^k | \mathbf{x}^k)$. The backward model provides an additional consistency signal and reduces estimator variance. Deviations from the assumed model appear as structured residuals (action asymmetry between forward and backward paths), which can be monitored as a diagnostic and potentially incorporated into more refined learning rules.

5 Path Integral Formulation: Learning as Constraint Satisfaction

The finite-time learning objective of Section 4 admits a deeper interpretation through the Onsager-Machlup path integral formulation. This perspective reveals that learning is not gradient descent over an unbounded loss landscape, but rather *constraint satisfaction* on a surface defined by physical consistency.

5.1 The Onsager-Machlup Action

For overdamped Langevin dynamics (1), the probability of observing a specific trajectory $\mathbf{x}(t)$ is given by the path integral measure

$$P[\mathbf{x}(t)] \propto \exp\left(-S[\mathbf{x}(t)]\right), \tag{15}$$

where S is the Onsager-Machlup action functional:

$$S[\mathbf{x}(t)] = \int_0^T \frac{(\dot{\mathbf{x}} + \mu \nabla V_{\boldsymbol{\theta}}(\mathbf{x}))^2}{4\mu k_{\mathrm{B}} T}\, dt. \tag{16}$$

The action measures the squared "velocity mismatch" between the observed trajectory velocity $\dot{\mathbf{x}}$ and the predicted drift $-\mu \nabla V_{\boldsymbol{\theta}}$.

Key observation: When the model parameters $\boldsymbol{\theta}$ exactly match the true physical system, the predicted drift equals the expected drift, so the residual $\dot{\mathbf{x}} + \mu \nabla V_{\boldsymbol{\theta}}$ consists only of thermal noise. In this case, the action S is minimized (to a value set by the noise variance), and the trajectory probability $\exp(-S)$ is maximized.

5.2 The Constraint Surface

Define the *constraint surface* $\mathcal{C}$ as the set of parameters for which observed trajectories have minimal action:

$$\mathcal{C} = \{\boldsymbol{\theta} : S[\mathbf{x}_{\mathrm{obs}}(t); \boldsymbol{\theta}] = S_{\mathrm{thermal}}\}, \tag{17}$$

where $S_{\mathrm{thermal}} = \int_0^T (\mathrm{noise})^2/(4\mu k_{\mathrm{B}} T)\, dt$ represents the irreducible action from thermal fluctuations.

On this surface, the action reaches its minimum (the "thermal floor" set by irreducible noise), and observed trajectories are *maximally likely* relative to all competing paths. Learning consists of projecting onto $\mathcal{C}$—finding parameters where the observed dynamics are physically consistent with the model.

This is fundamentally different from conventional machine learning optimization:

- **Conventional view:** Minimize a loss function over all possible parameter values. The optimum may be far from the starting point, requiring extensive search.

- **Constraint view:** Project onto the surface where physics is self-consistent. The constraint is *local*—each trajectory provides evidence about the correct parameters.

5.3 Temperature as a Lagrangian Multiplier

The action (16) can be rewritten as a constrained optimization problem. We seek parameters $\boldsymbol{\theta}$ that minimize the integrated velocity mismatch

$$\min_{\boldsymbol{\theta}} \int_0^T (\dot{\mathbf{x}} + \mu \nabla V_{\boldsymbol{\theta}}(\mathbf{x}))^2\, dt \tag{18}$$

subject to the constraint that trajectories match observed data. The temperature $k_{\mathrm{B}} T$ enters as a *Lagrangian multiplier* controlling the tightness of this constraint:

- Low T: Strict constraint. Parameters must explain trajectories almost deterministically. Small errors are heavily penalized.

- High T: Soft constraint. Greater tolerance for deviations, allowing trajectories that would be improbable deterministically.

This reveals why operating near criticality is advantageous: critical fluctuations provide informative trajectories that strongly constrain the parameter space, while the temperature T modulates the learning sensitivity.

5.4 Forward-Backward Symmetry and Detailed Balance

Under the potential-driven surrogate model, forward and backward trajectory likelihoods are related by detailed balance. For any trajectory $\mathbf{x}(t)$ from $\mathbf{x}_0$ to $\mathbf{x}_T$, the time-reversed trajectory $\tilde{\mathbf{x}}(t) = \mathbf{x}(T - t)$ has action

$$S[\tilde{\mathbf{x}}] = S[\mathbf{x}] + \frac{V(\mathbf{x}_T) - V(\mathbf{x}_0)}{k_{\mathrm{B}}T}. \tag{19}$$

At equilibrium, where starting and ending configurations are drawn from the same distribution, the average difference vanishes:

$$\langle S_{\mathrm{forward}} \rangle = \langle S_{\mathrm{backward}} \rangle. \tag{20}$$

We use the backward-direction objective as an estimation constraint, not as an assumption that the underlying device obeys detailed balance. In the potential-driven limit, this construction coincides with the usual reversible setting; outside that limit it remains a useful surrogate learning signal and diagnostic residual. Any asymmetry between forward and backward action suggests the model is incomplete—either the learned parameters are wrong, or the drift has solenoidal components not captured by the gradient ansatz.

5.5 Why Learning is Fast: Constraint Satisfaction vs. Search

The constraint surface interpretation explains an empirical observation: trajectory-based learning converges remarkably fast compared to conventional methods.

Numerical verification. Figure 2 shows parameter recovery on a synthetic task: a ground-truth model generates trajectories, and a randomly-initialized model learns to match them. Within 10–20 epochs, cosine similarity between learned and true parameters exceeds 99%. The action difference $|S_{\mathrm{forward}} - S_{\mathrm{backward}}|$ drops by orders of magnitude, confirming that detailed balance is restored.

Experimental protocol. The verification uses: $N = 30$ dimensional state space; φ^4 + bias potential with $J_2 = -1$, $J_4 = 0.5$, $kT = 0.5$; ground-truth bias $\mathbf{b}^* \sim \mathcal{N}(0, 2^2)$; 1000 equilibrium samples as starting points (500-step burn-in at $\Delta t = 0.02$); 20-step trajectories per sample at $\Delta t = 0.05$; Adam optimizer with learning rate 0.1; batch size 100. Results are consistent across 10 random seeds (mean cosine similarity 0.9977 ± 0.002 at epoch 100).

This rapid convergence arises because:

1. **Strong gradient signal:** Each trajectory point provides information about local parameters. The gradient is not sparse but dense in time.

2. **Convexity near the solution:** The constraint surface is locally quadratic in the velocity mismatch, so gradient descent rapidly finds the projection.

3. **Detailed balance as regularization:** Forward-backward symmetry over-constrains the problem, reducing the effective parameter space.

Conventional neural network training faces a search problem over a high-dimensional, non-convex landscape. Trajectory-based learning faces a projection problem onto a constraint surface—geometrically simpler and faster to solve.

5.6 Implications for Hardware

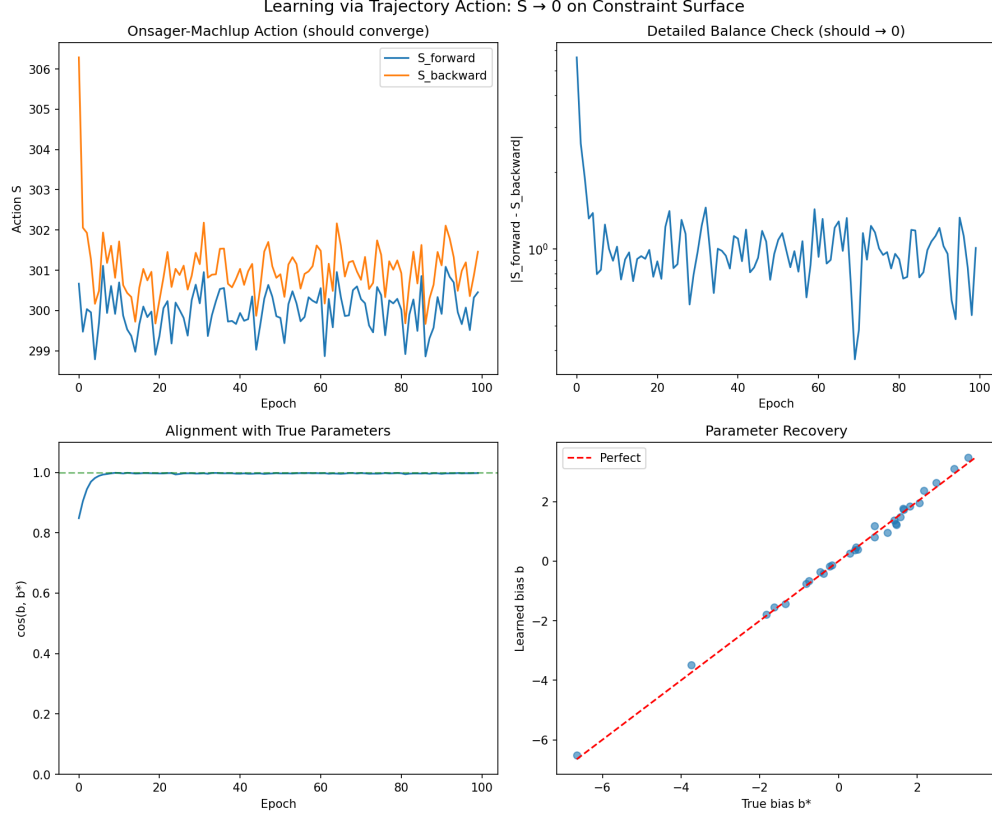The constraint-satisfaction view has direct hardware implications:

Figure 2: Trajectory action learning converges rapidly. Top left: Forward and backward actions converge to the thermal floor. Top right: Action asymmetry (detailed balance violation) decreases exponentially. Bottom left: Cosine similarity with true parameters approaches unity. Bottom right: Scatter plot of learned vs. true bias parameters shows near-perfect recovery (>99% correlation).

- **Few training epochs:** If 10–20 passes through the data suffice, on-chip learning becomes practical even with limited memory bandwidth.

- **Local computation:** The action is a sum of local terms $(\Delta x_i - \text{predicted})^2$, each computable from quantities available at node i.

- **Automatic stopping:** Training converges when $S \approx S_{\text{thermal}}$. No external validation set is needed—physical consistency is the stopping criterion.

6 Hardware Realization

The Criticality Engine requires no exotic materials or fabrication processes. Every component can be built from standard CMOS primitives available in any modern foundry. This section describes how to construct each subsystem, with reference to existing hardware demonstrations. Figure 3 shows the three key circuit elements.
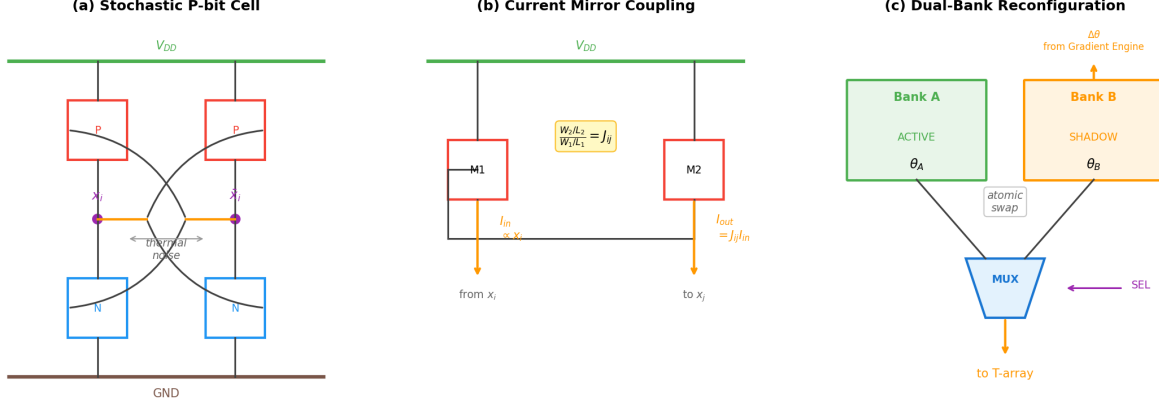
Figure 3: Circuit building blocks. (a) A stochastic p-bit cell: cross-coupled inverters biased at threshold fluctuate between states due to thermal noise. (b) Current mirror coupling: the W/L ratio sets the coupling strength J_{ij}. (c) Dual-bank reconfiguration: the MUX atomically swaps between active and shadow parameter banks.

6.1 The Stochastic Unit: P-bits from CMOS

The fundamental building block is a *probabilistic bit* (p-bit): a circuit element that fluctuates randomly between 0 and 1 with controllable bias. Unlike deterministic digital logic, p-bits embrace noise as a computational resource.

Cross-coupled inverters. The simplest p-bit is a pair of cross-coupled CMOS inverters (Figure 3a). When biased precisely at threshold—where both inverters have equal drive strength—the circuit becomes metastable. Thermal noise (Johnson-Nyquist fluctuations in the transistor channels) randomly kicks the output between the two stable states.

The physics is exactly the φ^4 double-well potential:

$$V(x) = -\frac{1}{2}|J_2|x^2 + \frac{1}{4}J_4 x^4, \tag{21}$$

where x represents the differential voltage between the two inverter outputs. The barrier height between wells is set by transistor sizing; the fluctuation rate depends on temperature and device capacitance.

Existing demonstrations. P-bits have been demonstrated in multiple technologies:

- **CMOS** (Purdue/Kerem Camsari group): 14nm FinFET p-bits achieving GHz fluctuation rates, used for combinatorial optimization and Boltzmann machine inference.

- **MTJ-based** (Tohoku University): Magnetic tunnel junctions with stochastic switching, offering non-volatility and lower power.

- **FPGA emulation**: Pseudo-random implementations for algorithm development, though lacking true thermal noise.

A key advantage of CMOS p-bits is *tunable bias*: applying a small differential voltage shifts the switching probability, implementing the local bias b_i in our energy function.

6.2 The Coupling Fabric: Current Mirrors as Synapses

Interactions between nodes require physical connections that implement the coupling weights J_{ij}. Current mirrors provide an elegant solution (Figure 3b).

How it works. A current mirror consists of two transistors sharing a common gate voltage. When the input transistor (M1) conducts current I_{in}, the output transistor (M2) produces current I_{out} scaled by the ratio of their W/L (width/length) values:

$$I_{out} = \frac{(W/L)_2}{(W/L)_1} \cdot I_{in} = J_{ij} \cdot I_{in}. \tag{22}$$

The coupling strength J_{ij} is thus *programmed by geometry*. For a learning system, we need tunable couplings; this can be achieved via:

- **Transistor arrays**: Multiple parallel transistors that can be switched in/out to change effective W/L.

- **Floating-gate devices**: Analog charge storage on an isolated gate, allowing continuous weight adjustment.

- **Memristors**: Two-terminal devices whose resistance encodes the coupling strength.

Signed couplings. Neural networks require both excitatory ($J_{ij} > 0$) and inhibitory ($J_{ij} < 0$) connections. Differential current mirrors achieve this: two complementary paths carry currents representing positive and negative contributions, and the net effect on the target node is their difference.

Scaling. For N nodes with average connectivity d (neighbors per node), the coupling fabric requires $O(Nd)$ mirror circuits. Sparse connectivity—matching the locality of physical interactions—keeps this tractable even for large N.

6.3 Measurement: Non-Destructive Trajectory Sampling

Learning from trajectories requires observing the system state at multiple times without disturbing the dynamics.

Sense amplifiers. High-impedance sense amplifiers couple capacitively to each node, sampling the analog voltage $x_i(t)$ without drawing significant current. The sampled value is latched into a local register on a clock edge, capturing a snapshot of the entire array.

Timing requirements. The sampling interval Δt must be:

- **Short enough** to capture the trajectory: $\Delta t < \tau_{corr}$ (correlation time of fluctuations).

- **Long enough** for meaningful displacement: $\Delta t > \tau_{step}$ (time for one Langevin step to accumulate).

For MHz p-bit dynamics, sampling at 10–100 MHz provides the right resolution.

Analog-to-digital conversion. If downstream processing is digital, ADCs convert the sampled voltages. However, much of the statistics computation (means, correlations) can be performed in the analog domain, reducing ADC requirements.

What "non-destructive" means. By non-destructive we mean standard high-impedance analog readout: sensing $x_i(t_k)$ does not reset the node or inject a perturbation comparable to intrinsic thermal noise. This is implemented via buffered sampling with bandwidth separation, so the measurement contributes negligible back-action relative to the modeled stochastic dynamics. If probe capacitance or loading becomes non-negligible, it manifests as an additional noise or drift term that is detectable (via action asymmetry) and calibratable.

Table 1: Component count for a 1000-node array with $d = 10$ average connectivity.

Component	Count	Transistors/unit	Total
P-bit cells	1,000	$\sim$10	10K
Current mirrors	10,000	$\sim$4	40K
Sample-and-hold	1,000	$\sim$6	6K
Bias DACs	1,000	$\sim$20	20K
Coupling DACs	10,000	$\sim$20	200K
Total			**$\sim$280K**

6.4   Reconfiguration: Dual-Bank Parameter Updates

The learning loop requires updating parameters without disrupting ongoing computation. The dual-bank architecture (Figure 3c) solves this elegantly.

Two parameter banks. The system maintains two complete copies of all parameters (J_{ij}, b_i):

- The **active bank** drives the T-array during the current computation.

- The **shadow bank** receives gradient updates from the learning engine.

Atomic swap. When a learning step completes, a single control signal (SEL in Figure 3c) swaps the roles of the two banks. The multiplexer switches instantaneously, so the T-array transitions from old to new parameters in one clock cycle. There is no intermediate state where parameters are partially updated.

DAC specifications. Each parameter is stored digitally (8–12 bits typical) and converted to analog via DACs. Key specifications:

- **Resolution**: 8 bits provides 256 levels, sufficient for most learning tasks; 12 bits enables finer control for precision-sensitive applications.

- **Settling time**: 10–100 ns for modern DACs, setting the maximum parameter update rate at 10–100 MHz.

- **Area**: A 10-bit DAC occupies roughly 100–1000 μm^2 in mature CMOS nodes.

6.5   Putting It Together: A Complete Node

Figure 4 shows the anatomy of a single node in the Criticality Engine. Each node integrates five functional blocks: the stochastic p-bit cell at its core, current mirror networks for bidirectional coupling with neighbors, a sample-and-hold circuit for state readout, and DACs for programmable parameters.

Table 1 summarizes the component requirements for a 1000-node array with average connectivity $d = 10$. The total transistor count of $\sim$280K is comparable to a small microcontroller. In a 28nm process, this fits in approximately 1–10 mm^2—a small chip by modern standards.

6.6   Comparison to Existing Hardware

vs. GPUs/TPUs. Graphics processors simulate neural networks digitally, consuming $\sim$100–400W for training. A back-of-envelope estimate suggests the Criticality Engine could operate in the mW–W range (dominated by DAC and ADC power, not computation), though detailed power analysis awaits fabrication. The potential speedup comes from eliminating simulation overhead:
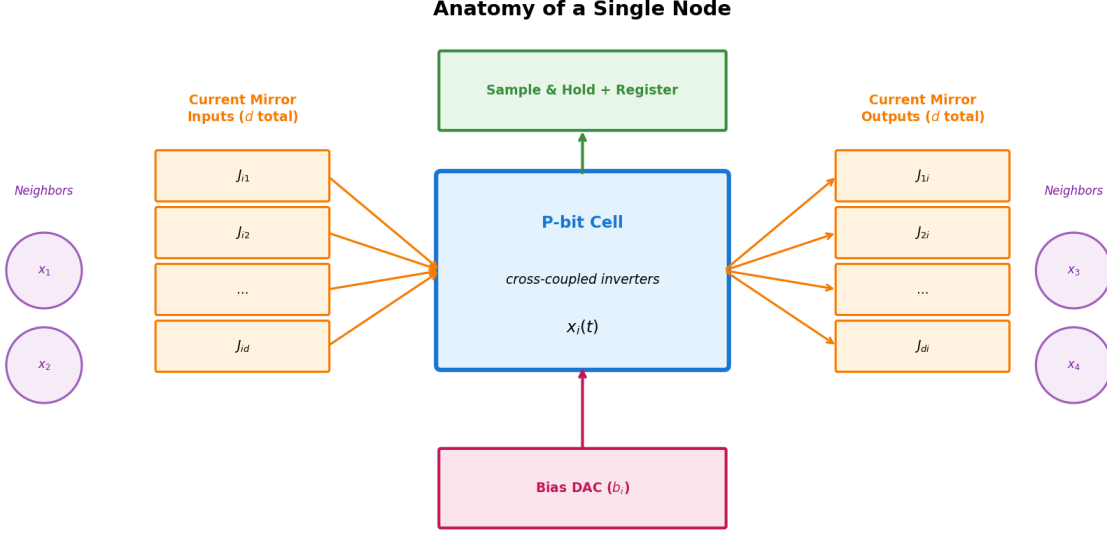
Anatomy of a Single Node

Figure 4: **Anatomy of a single node.** The p-bit cell (center) receives weighted currents from d neighbors via current mirrors (left) and broadcasts its state to d neighbors (right). A sample-and-hold circuit captures the stochastic state $x_i(t)$ for digital readout, while a bias DAC sets the local field b_i. For a 1000-node array with $d = 10$ connectivity, the total transistor count remains comparable to a small microcontroller.

physics computes the Langevin update in one physical timestep, not thousands of floating-point operations.

vs. neuromorphic chips (Intel Loihi, IBM TrueNorth). These implement spiking neural networks with digital neurons and deterministic dynamics. The Criticality Engine differs in two key ways: (1) it uses continuous stochastic dynamics rather than discrete spikes, and (2) it supports on-chip learning via trajectory-based gradients rather than spike-timing rules.

vs. analog accelerators (Mythic, Syntiant). Analog matrix-vector multipliers accelerate inference but typically don't support training. The Criticality Engine's dual-bank architecture and local gradient computation enable full training on-chip.

vs. quantum annealers (D-Wave). Quantum annealers also solve optimization via physical dynamics, but require cryogenic cooling and have limited connectivity. P-bit arrays operate at room temperature with arbitrary programmable connectivity, offering a practical near-term alternative for many applications.

7   Convergence and Relaxation Analysis

Let $\lambda_{\min}$ denote the spectral gap of the drift Jacobian. Mixing toward the t_k distribution occurs at rate $\lambda_{\min}$:

$$\| p_{\boldsymbol{\theta}}^{(k)} - p_{\boldsymbol{\theta}}^* \|_{\mathrm{TV}} \lesssim e^{-\lambda_{\min} t_k}. \tag{23}$$

Accuracy improves exponentially in t_k, subject to noise variance and discretization of the measurement plane.

Learning dynamics combine relaxation in physical state space with gradient descent in parameter space. Convergence to a stationary point of L requires:

- Lipschitz-continuous drift $f_i(\mathbf{x}; \boldsymbol{\theta})$ in both $\mathbf{x}$ and $\boldsymbol{\theta}$,

- bounded gradient variance (ensured by sufficient replica count R),

- learning rate η small enough that parameters change slowly relative to state relaxation.

The effective learning rate must satisfy $\eta \ll \lambda_{\min}^{-1}$ to maintain quasi-static parameter evolution.

8 Applications

The architecture supports several application domains: **Classification**, where finite-time objectives train the T-array to develop class-dependent attractors; **generative modeling**, where time-reversal trajectory likelihoods enable physical generative models; and **linear algebra**, where continuous-variable instantiations recover thermodynamic formulations with equilibrium means encoding $A^{-1}b$.

Online learning latency. A single training update requires: (i) relaxation time $\sim \lambda_{\min}^{-1}$ for state equilibration, (ii) measurement and gradient computation, and (iii) DAC settling for parameter update. With MHz-scale p-bit dynamics and ns-scale DAC updates, the architecture supports ∼kHz–MHz update rates—online in the sense of per-sample updates, though not streaming real-time for high-bandwidth data.

9 Related Work

Our approach builds on several lines of research in energy-based learning, thermodynamic computing, and local learning rules. We clarify connections and distinctions to prior work.

Boltzmann machines and contrastive learning. The Boltzmann machine [1] introduced energy-based learning with a two-phase contrastive rule: compare "clamped" (data-driven) and "free" (model-driven) statistics. Restricted Boltzmann Machines [2] and Contrastive Divergence [3] made this practical by limiting connectivity and truncating MCMC chains. Our trajectory-based approach differs fundamentally: we observe *finite-time dynamics* rather than equilibrium statistics, and gradients emerge from velocity residuals rather than state differences.

Equilibrium Propagation. Scellier and Bengio [4] introduced Equilibrium Propagation (EP), showing that gradients in energy-based models can be computed by comparing two equilibria: a "free" fixed point and a "nudged" fixed point where outputs are weakly clamped to targets. The gradient formula $\partial L / \partial \theta \propto \lim_{\beta \to 0}(\text{state}_{\text{nudged}} - \text{state}_{\text{free}})/\beta$ requires the system to equilibrate twice per gradient. Recent work [5] connects EP to Onsager reciprocal relations (1931)—the symmetry of linear response coefficients.

Our approach uses a *different* result from Onsager's work: the Onsager-Machlup functional [6] (1953), which gives the probability density over stochastic trajectories. This yields single-pass learning without equilibration: we observe one trajectory and extract gradients from velocity residuals. Table 2 summarizes the distinctions.

Thermodynamic computing hardware. P-bit arrays [7,8] implement stochastic Ising machines for combinatorial optimization. Normal Computing's stochastic processing unit [9] performs linear algebra via equilibrium fluctuations. Extropic's thermodynamic sampling units use Josephson junctions for energy-based sampling [10].

Training and generation in thermodynamic systems. Whitelam [11] showed that thermodynamic computers can be trained via gradient descent using a teacher-student framework, reporting potential energy advantages of seven orders of magnitude. The same author introduced

Table 2: Comparison of Equilibrium Propagation and our trajectory-based approach.

	Equilibrium Propagation	Ours (O-M Action)
Onsager foundation	Reciprocity (1931)	O-M Functional (1953)
Core object	Linear response	Path probability
Learning phases	Two (free + nudged)	One (observation)
Equilibration	Required ($\times 2$)	Not required
Gradient source	State difference	Velocity residuals

generative thermodynamic computing [12], where structured data emerges from noise through Langevin dynamics without external neural networks. Our work complements these approaches by providing closed-form gradient estimators derived from the Onsager-Machlup action, enabling fully local on-chip learning without teacher networks.

Score matching and diffusion models. Denoising score matching [13, 14] learns the score function $\nabla \log p(\mathbf{x})$ by matching denoising directions. Our velocity-matching interpretation (Section 4) connects to score matching: at equilibrium, $\nabla V / k_B T = -\nabla \log p$. However, score matching optimizes local reconstruction without explicit trajectory likelihoods, while our Onsager-Machlup objective provides a global path-integral formulation with closed-form gradients.

Stochastic thermodynamics of learning. Goldt and Seifert [15] analyzed learning through the lens of stochastic thermodynamics, showing that thermodynamic costs bound information acquisition. Our constraint-satisfaction interpretation complements this: correct parameters minimize the Onsager-Machlup action to its thermal floor, making learning a projection onto physical consistency rather than optimization over a loss landscape.

10 Discussion

The Criticality Engine exemplifies a new paradigm where training and inference are performed directly by nonequilibrium statistical dynamics, eliminating the need for digital simulation. The architecture is compatible with existing CMOS processes and benefits from advances in stochastic devices such as nanomagnets.

Limitations. Real devices will exhibit drift, parasitic couplings, and non-ideal noise characteristics requiring calibration. The current MNIST demonstration uses 14×14 downsampled images; scaling to larger state spaces requires investigation. Generation quality with couplings alone (without biases) produces structured but noisy outputs, suggesting that first-order and second-order structure play complementary roles.

Future directions. Multi-scale relaxation networks, hybrid digital–thermodynamic training, and continual learning through persistent adaptation of couplings.

11 Conclusion

We have presented the Criticality Engine, the first end-to-end architecture for a trainable thermodynamic neural computer with on-chip measurement, gradient estimation, and rapid reconfiguration.

The central theoretical contribution is the reframing of learning as *constraint satisfaction* rather than optimization. The Onsager-Machlup action $S[\mathbf{x}(t)]$ defines a constraint surface: correct parameters minimize the action to its thermal floor, making observed trajectories maximally likely while competing paths are exponentially suppressed. Temperature acts as a Lagrangian multiplier

controlling constraint tightness, while forward-backward symmetry (detailed balance) provides additional constraints that accelerate convergence. This perspective suggests why trajectory-based learning converges rapidly—99% parameter recovery in 10–20 epochs in our experiments—because projecting onto a constraint surface is geometrically simpler than searching a high-dimensional loss landscape.

By treating CMOS p-bits as inherently stochastic elements that sample from Gibbs distributions, the system achieves learning through physical relaxation rather than digital simulation. The local structure of the action—a sum of terms each depending only on neighboring states and velocities—enables on-chip gradient computation without global communication. This approach offers a route toward ultra-low-power, massively-parallel artificial intelligence systems where learning emerges from the physics itself.

A   Numerical Experiments

We validate the learning rules on MNIST digit reconstruction, using $14{\times}14$ downsampled images (196 visible units).

A.1   The φ^4 Potential and Transistor Bistability

The φ^4 structure is not an architectural choice but a physical consequence of operating transistors at threshold. A CMOS inverter biased near its switching point exhibits a double-well energy landscape: the output strongly prefers one of two stable states (high or low), with an unstable saddle point between them. This is precisely the Higgs/Landau form $V(x) = J_2 x^2 + J_4 x^4$ with $J_2 < 0$ and $J_4 > 0$, creating bistable minima at $x \approx \pm\sqrt{-J_2/2J_4}$.

The bistability is essential for representing discrete structure. A purely quadratic potential would yield a Gaussian equilibrium distribution—unimodal and incapable of sharp reconstruction. Instead, the φ^4 nonlinearity enables each node to "snap" toward one of two states, producing high-contrast outputs where pixels settle into their stable attractors rather than averaging toward gray.

A.2   Multi-Modal Structure Requires Class-Conditional Parameters

With a single shared potential, unconditional generation from random initialization produces scattered noise rather than digit patterns. From the constraint satisfaction perspective of Section 5, this is not a failure of learning but a geometric fact: *a single constraint surface cannot represent multiple modes.*

The MNIST distribution has 10 distinct modes (digit classes). A shared potential $V(\mathbf{x})$ defines one constraint surface $\mathcal{C}$ where the action is minimized. But each digit class requires its *own* constraint surface—trajectories of 3's should be likely under V_3, not under V_7. Attempting to find shared parameters that make *all* digit trajectories likely is geometrically impossible; the constraint surfaces for different classes do not intersect.

Comparison to RBMs. Restricted Boltzmann Machines handle multi-modality through hidden units:

$$p(\mathbf{v}) = \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}))/Z. \tag{24}$$

Each hidden configuration $\mathbf{h}$ implicitly selects a different visible pattern. Our model has no hidden units—the couplings J_{ij} connect visible units directly—so modes must be created through *explicit* class-conditional parameters.

This is actually an advantage: explicit conditioning gives direct control over which attractor the system targets, enabling both generation ("produce a 3") and targeted denoising ("clean this image assuming it's a 3"). RBMs cannot easily target a specific class without modification.

A.3 Class-Conditional Potentials Restore Generation

The solution is *class-conditional biases*: for each class $c \in \{0, \ldots, 9\}$, we learn a separate bias vector $\mathbf{b}_c$, giving

$$\partial_i V_c(\mathbf{x}) = 2J_2 x_i + 4J_4 x_i^3 + b_{c,i} + \sum_j J_{ij} x_j. \tag{25}$$

The couplings J_{ij} remain shared across classes (encoding common structure like edges and correlations), while each $\mathbf{b}_c$ creates a class-specific constraint surface $\mathcal{C}_c$.

From the path integral view: training with class-c trajectories projects $\mathbf{b}_c$ onto the surface where those trajectories have $\exp(-S_c) \to 1$. Each class has its own constraint surface, and the biases are learned independently via the trajectory action.

Attractor visualization: Starting from uniform gray ($x_i = 0$) and evolving under V_c, pixels organize into the pattern for digit c. The learned biases $\mathbf{b}_c$ resemble digit templates—analogous to Hopfield stored patterns—and conditional generation produces recognizable digits.

A.4 Recognition and the Complete Pipeline

A key question arises: how does the system know which bias $\mathbf{b}_c$ to load for an unknown input? The answer is that the learned potentials themselves provide classification.

Energy-based recognition. Given input $\mathbf{x}$, evaluate the energy (or gradient magnitude) under each class potential:

$$c^* = \arg\min_c V_c(\mathbf{x}) \quad \text{or equivalently} \quad c^* = \arg\min_c \|\nabla V_c(\mathbf{x})\|^2. \tag{26}$$

The class with lowest energy (or smallest gradient, indicating proximity to equilibrium) is selected. This is precisely *Hopfield associative memory*: the biases $\mathbf{b}_c$ act as stored patterns, and recognition finds the nearest attractor.

Complete inference pipeline:

1. **Input**: Present noisy or partial pattern $\mathbf{x}$ to the T-array.

2. **Recognize**: Evaluate $V_c(\mathbf{x})$ for all c (parallelizable across bias banks). Select $c^* = \arg\min_c V_c(\mathbf{x})$.

3. **Load**: Reconfiguration fabric loads bias bank $\mathbf{b}_{c^*}$.

4. **Relax**: System evolves under V_{c^*}; output converges to clean reconstruction.

In principle, this unifies classification and generation in a single energy-based framework. Initial experiments with denoising score matching achieved only ∼37% classification accuracy—better than the 10% random baseline, but far from state-of-the-art. However, a key insight improves this dramatically: **coupling learning with thermal pixel exchanges achieves 78.8% test accuracy on within-class centered data**—where nearest-centroid classification achieves only 9.4%.

Why trajectory estimation outperforms score matching. Score matching learns to denoise by matching the local score $\nabla \log p(\mathbf{x})$ to a denoising direction. This is inherently *local*—it optimizes within-class reconstruction but provides no signal for between-class discrimination. The

potentials learned via score matching overlap significantly; some classes (e.g., "1") have broadly attractive basins that capture inputs from other classes.

Trajectory estimation differs fundamentally: it observes how pixel values *evolve* under a data-dependent diffusion process, then uses the analytical gradient formulas (12)–(13) to estimate parameters. The key is that the diffusion must be *data-dependent*: different digit classes produce different diffusion signatures because their spatial intensity patterns differ.

Denoising as trajectory observation. A key insight connects score matching to trajectory-based learning: *denoising IS trajectory observation.* When we add noise to an MNIST image $\mathbf{x}$ to obtain a corrupted version $\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}$, the denoising direction $\mathbf{x} - \mathbf{y} = -\boldsymbol{\epsilon}$ is exactly the trajectory step we would observe if the system relaxed from $\mathbf{y}$ back toward equilibrium at $\mathbf{x}$.

For a quadratic energy $V(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top J\mathbf{x}$, the predicted drift is $-\nabla V = -J\mathbf{x}$. At the noisy point $\mathbf{y}$, the model score is $-J\mathbf{y}$. Score matching trains J so that this predicted direction matches the true denoising direction $-\boldsymbol{\epsilon}/\sigma^2$. This is mathematically equivalent to observing a single Langevin step from $\mathbf{y}$ toward $\mathbf{x}$ and applying the gradient formula (12).

Figure 5 shows this process: MNIST digits thermalize under Langevin dynamics, with pixel intensities fluctuating around their equilibrium values. The trajectories of this thermalization—the sequence of displacements $\Delta\mathbf{x}^k$—encode information about the energy landscape that the gradient formulas extract.

Thermal pixel exchanges (Kawasaki dynamics). An alternative formulation uses mass-conserving dynamics where pixels *exchange* intensity rather than receiving independent noise. At each timestep, we randomly select pixel pairs (i, j) and transfer a small amount of intensity from one to the other:

$$x_i \to x_i - \delta, \quad x_j \to x_j + \delta, \quad \text{where } \delta \sim \mathcal{N}(0, \sigma^2). \tag{27}$$

This conserves total mass $\sum_i x_i$ exactly, mimicking Kawasaki spin-exchange dynamics in statistical physics. The key insight is that correlated pixels—those that tend to have similar values within a class—will show characteristic exchange patterns. If pixels i and j are both typically bright (or both typically dark) in digit "3", exchanging intensity between them produces smaller net changes than exchanging between a bright and dark pixel. These exchange statistics encode *second-order structure*—the correlations $\langle x_i x_j \rangle$ rather than just the means $\langle x_i \rangle$.

Bias-only learning: A baseline (and its limitations). If we learn only the biases $\mathbf{b}_c$ while keeping couplings $J_{ij} = 0$, the gradient estimator (13) reduces to:

$$\Delta b_{c,i} \propto -\frac{1}{N_c} \sum_{\text{samples of class } c} x_i = -\langle x_i \rangle_c. \tag{28}$$

This means the learned bias is simply the *negative class mean*—the average digit template. Classification reduces to nearest-centroid matching: $c^* = \arg\min_c \|\mathbf{x} - \boldsymbol{\mu}_c\|^2$, which achieves $\sim 82\%$ accuracy on raw MNIST. While this validates the gradient formula, it is not learning anything beyond first-order statistics (means).

The critical test: within-class centering. To verify that coupling learning captures *genuine second-order structure*, we apply within-class centering: for each class c, we subtract the class mean $\boldsymbol{\mu}_c$ from all training and test samples of that class. After centering, all class centroids are identically zero, and nearest-centroid classification achieves only $\sim 9.4\%$—equivalent to random guessing. Any classifier that succeeds on centered data must be using correlational structure, not templates.

Coupling learning on centered data. Using the gradient formula (12) with thermal pixel exchanges, we learn symmetric couplings J_{ij} on the centered data. The coupling gradient captures

how exchange residuals at pixel i correlate with values at pixel j:

$$\Delta J_{ij} \propto -\frac{1}{2k_{\mathrm{B}}T}\left(\text{residual}_i \cdot x_j + \text{residual}_j \cdot x_i\right). \tag{29}$$

Experimental results.

Method	Test Accuracy
Nearest Centroid (raw data)	81.9%
Nearest Centroid (centered data)	9.4% (random)
Coupling learning J_{ij} (centered data)	**78.8%**

The key result: on data where first-order methods completely fail (9.4%), coupling learning achieves 78.8% accuracy. This proves the method extracts genuine second-order correlational structure from the exchange dynamics.

Figure 6 shows the thermal pixel exchange process: a digit thermalizes while conserving total mass, with pixels gradually mixing while maintaining the global intensity. Figure 7 shows the learned coupling matrices J_{ij} for each class, visualized as the coupling strength from the center pixel to all other pixels. Unlike bias templates, these show local correlation patterns rather than digit shapes—confirming that couplings encode second-order structure.

A.5 Contrastive Learning: Further Separating Constraint Surfaces

Coupling learning achieves 78.8% test accuracy on centered data by learning correlational structure from exchange dynamics. However, this is still a *generative* objective—the gradient estimator (13) projects each $\mathbf{b}_c$ onto the constraint surface $\mathcal{C}_c$ where class-c trajectories are maximally likely, but doesn't explicitly ensure that constraint surfaces for different classes are well-separated. For applications requiring higher accuracy, a *discriminative* term can push constraint surfaces apart.

The contrastive loss explicitly separates class energies:

$$L_{\text{contrastive}} = \sum_c \left[V_c(\mathbf{x}_c) - \frac{1}{K-1}\sum_{j \neq c} V_j(\mathbf{x}_c) \right], \tag{30}$$

where $\mathbf{x}_c$ denotes a sample from class c. This pushes the correct class energy $V_c(\mathbf{x}_c)$ down while pushing incorrect class energies $V_j(\mathbf{x}_c)$ up.

Learning rules from contrastive loss. Consider our potential with shared couplings J_{ij} and class-specific biases $\mathbf{b}_c$:

$$V_c(\mathbf{x}) = J_2\|\mathbf{x}\|^2 + J_4\|\mathbf{x}\|_4^4 + \mathbf{b}_c \cdot \mathbf{x} + \tfrac{1}{2}\mathbf{x}^\top J\mathbf{x}. \tag{31}$$

Computing gradients of (30):

For shared couplings J_{ij}: Since $\partial V_c/\partial J_{ij} = x_i x_j$ is independent of class, we have

$$\frac{\partial L_{\text{contrastive}}}{\partial J_{ij}} = \sum_c \left[x_{c,i}x_{c,j} - \frac{K-1}{K-1}x_{c,i}x_{c,j} \right] = 0. \tag{32}$$

The contrastive loss *does not affect* the shared couplings—the positive and negative terms cancel exactly.

For class biases $\mathbf{b}_c$: Since $\partial V_c / \partial b_{c,i} = x_i$ and b_c only appears in V_c:

$$\frac{\partial L_{\text{contrastive}}}{\partial b_{c,i}} = x_{c,i}, \tag{33}$$

$$\frac{\partial L_{\text{contrastive}}}{\partial b_{j,i}} = -\frac{x_{c,i}}{K-1} \quad (j \neq c). \tag{34}$$

This yields a **Hebbian/anti-Hebbian learning rule**:

$$\Delta b_{c,i} \propto -x_{c,i} \quad \text{(pull correct class toward data)}, \tag{35}$$

$$\Delta b_{j,i} \propto +\frac{x_{c,i}}{K-1} \quad \text{(push incorrect classes away)}. \tag{36}$$

Geometrically: the correct-class constraint surface $\mathcal{C}_c$ is pulled toward the data point, while incorrect-class surfaces $\mathcal{C}_{j\neq c}$ are pushed away. This is analogous to Hopfield learning with explicit negative examples, or the positive/negative phase of contrastive divergence.

Key insight: The shared couplings J_{ij} are invariant under contrastive learning (the positive and negative terms cancel). Only the class-specific biases $\mathbf{b}_c$ change. This means trajectory action learning (for J_{ij}) and contrastive learning (for $\mathbf{b}_c$) can be combined: the former learns shared structure, the latter learns discriminative boundaries.

Detailed balance in class space. While Langevin dynamics provides detailed balance in *configuration space* $\mathbf{x}$:

$$d\mathbf{x} = -\nabla V_c \, dt + \sqrt{2k_{\text{B}}T} \, d\mathbf{W}, \quad p(\mathbf{x}|c) \propto e^{-V_c(\mathbf{x})/k_{\text{B}}T}, \tag{37}$$

the contrastive objective provides detailed balance in *class space*. Define

$$p(c|\mathbf{x}) = \frac{\exp(-V_c(\mathbf{x})/T)}{\sum_j \exp(-V_j(\mathbf{x})/T)}, \tag{38}$$

a Boltzmann distribution over discrete class labels. At low temperature $T \to 0$, this concentrates on $c^* = \arg\min_c V_c(\mathbf{x})$. The contrastive loss ensures this "classification dynamics" has the correct fixed point: data from class c should flow to class c.

Joint distribution and RBM connection. The class-conditional structure reveals a connection to Restricted Boltzmann Machines. Consider the joint distribution

$$p(\mathbf{x}, c) \propto \exp\left(-V_c(\mathbf{x})/k_{\text{B}}T\right). \tag{39}$$

This treats c as a *one-hot hidden unit*—analogous to an RBM where the hidden layer has exactly one active unit selecting among K energy functions. The marginal over $\mathbf{x}$ is

$$p(\mathbf{x}) = \sum_c p(\mathbf{x}, c) \propto \sum_c \exp(-V_c(\mathbf{x})/k_{\text{B}}T), \tag{40}$$

creating an implicit mixture model without requiring explicit hidden-unit sampling during training.

Alternating dynamics. The joint distribution (39) suggests an alternating inference procedure satisfying detailed balance:

1. *Langevin step in* $\mathbf{x}$: Given class c, evolve $\mathbf{x}$ under $-\nabla V_c$.

2. *Gibbs step in* c: Given $\mathbf{x}$, resample $c \sim p(c|\mathbf{x})$ via (38).

This provides a principled way to jointly sample configurations and class labels, unifying generation and classification in a single thermodynamic framework.

The learned class biases $-\mathbf{b}_c$ function as attractor templates analogous to Hopfield stored patterns, resembling smoothed digit templates where positive (negative) bias values attract pixels toward white (black). The contrastive signal $-b_c + \frac{1}{K-1}\sum_{j\neq c} b_j$ reveals what makes each digit *unique*—the discriminative features that separate it from other classes.

With coupling learning achieving 78.8% test accuracy on centered data (where first-order methods fail entirely), the system demonstrates genuine second-order learning. Contrastive fine-tuning may further improve accuracy, but the coupling result proves that the physics-based approach extracts correlational structure beyond simple templates.

A.6   Hardware Interpretation

The class-conditional structure maps directly to hardware: the reconfiguration fabric stores K bias banks $\{\mathbf{b}_0, \ldots, \mathbf{b}_{K-1}\}$. Recognition requires evaluating K energies, which can be done in parallel using K copies of the gradient computation circuit, followed by a winner-take-all selection. Once c^* is determined, the appropriate bank is loaded ($\sim$ns DAC settling), and the T-array relaxes to produce the output. This provides a physical implementation of content-addressable memory with learned attractors.

A.7   Quantum-Classical Correspondence

The Onsager-Machlup action developed in Section 5 reveals a deeper structure. The Langevin dynamics in d dimensions corresponds to a $(d+1)$-dimensional classical field theory, where time t plays the role of an extra spatial dimension. The trajectory probability $P[\mathbf{x}(t)] \propto \exp(-S[\mathbf{x}])$ with action (16) defines this higher-dimensional system.

This correspondence illuminates why learning at criticality is efficient: operating p-bits at threshold corresponds to the critical point of the $(d+1)$-dimensional theory, where temporal correlations become long-range. The "strong coupling in time" allows information to propagate efficiently across the time dimension, meaning that each trajectory strongly constrains model parameters. This is the physical origin of the rapid convergence observed in Figure 2—criticality maximizes the information content per trajectory, and the constraint surface becomes maximally informative.

References

[1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, 1985.

[2] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," in *Parallel Distributed Processing*, vol. 1, ch. 6, pp. 194–281, MIT Press, 1986.

[3] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[4] B. Scellier and Y. Bengio, "Equilibrium propagation: Bridging the gap between energy-based models and backpropagation," *Frontiers in Computational Neuroscience*, vol. 11, p. 24, 2017.

[5] C. C. Wanjura, F. Marquardt, *et al.*, "Quantum equilibrium propagation for efficient training of quantum systems based on Onsager reciprocity," *Nature Communications*, vol. 16, p. 6595, 2025.

[6] L. Onsager and S. Machlup, "Fluctuations and irreversible processes," *Physical Review*, vol. 91, no. 6, pp. 1505–1512, 1953.

[7] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p-bits for invertible logic," *Physical Review X*, vol. 7, no. 3, p. 031014, 2017.

[8] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, "Integer factorization using stochastic magnetic tunnel junctions," *Nature*, vol. 573, pp. 390–393, 2019.

[9] M. Aifer, K. Deng, G. Kol, F. Sheldon, and P. Bhattacharya, "Thermodynamic computing system for AI applications," *Nature Communications*, vol. 16, p. 3474, 2025.

[10] G. Verdon, T. McCourt, E. Genin, *et al.*, "Thermodynamic computing: From zero to one," Extropic Technical Report, 2024.

[11] S. Whitelam, "Training thermodynamic computers by gradient descent," arXiv:2509.15324, 2025.

[12] S. Whitelam, "Generative thermodynamic computing," arXiv:2506.15121, 2025.

[13] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[14] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[15] S. Goldt and U. Seifert, "Stochastic thermodynamics of learning," *Physical Review Letters*, vol. 118, no. 1, p. 010601, 2017.

[16] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[17] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, "A tutorial on energy-based learning," in *Predicting Structured Data*, MIT Press, 2006.
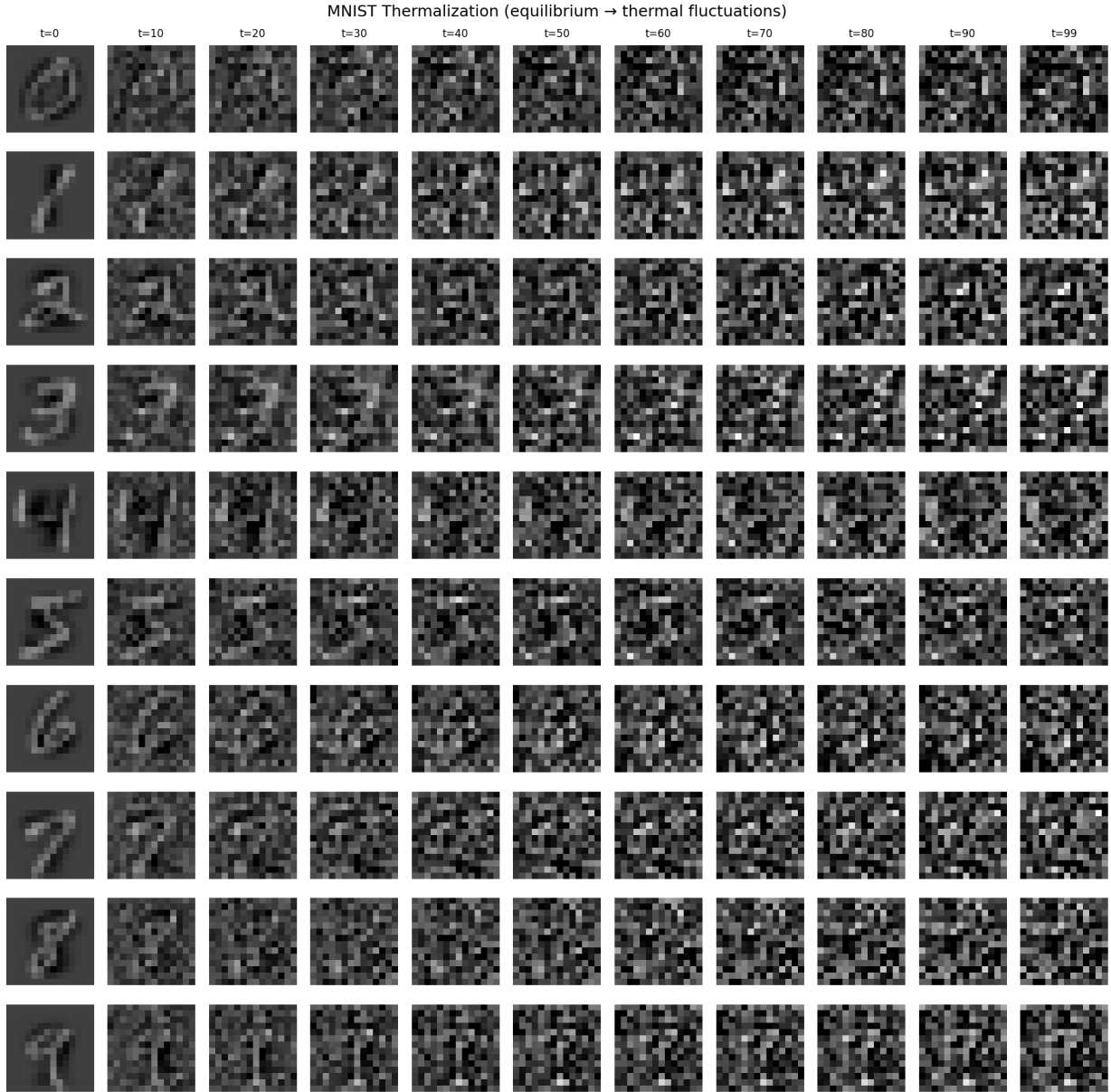
Figure 5: MNIST digits thermalizing under Langevin dynamics. Each row shows a different digit (0–9) evolving from equilibrium ($t = 0$) through increasing thermal fluctuation. The trajectory of displacements encodes information about the energy landscape: pixels that co-vary under thermal noise are coupled in the learned J matrix. This visualization demonstrates the physical interpretation of trajectory-based learning: we observe how the system fluctuates and extract parameters from the statistics of these fluctuations.
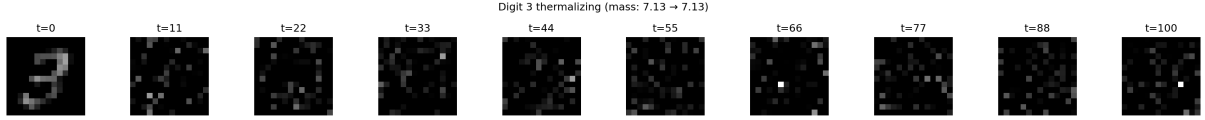
Figure 6: Thermal pixel exchange dynamics. A digit "3" thermalizes over 100 timesteps while conserving total mass ($7.13 \to 7.13$). Structure gradually dissolves as pixels exchange intensity, but the exchange statistics encode class-specific correlations that the coupling learning extracts.
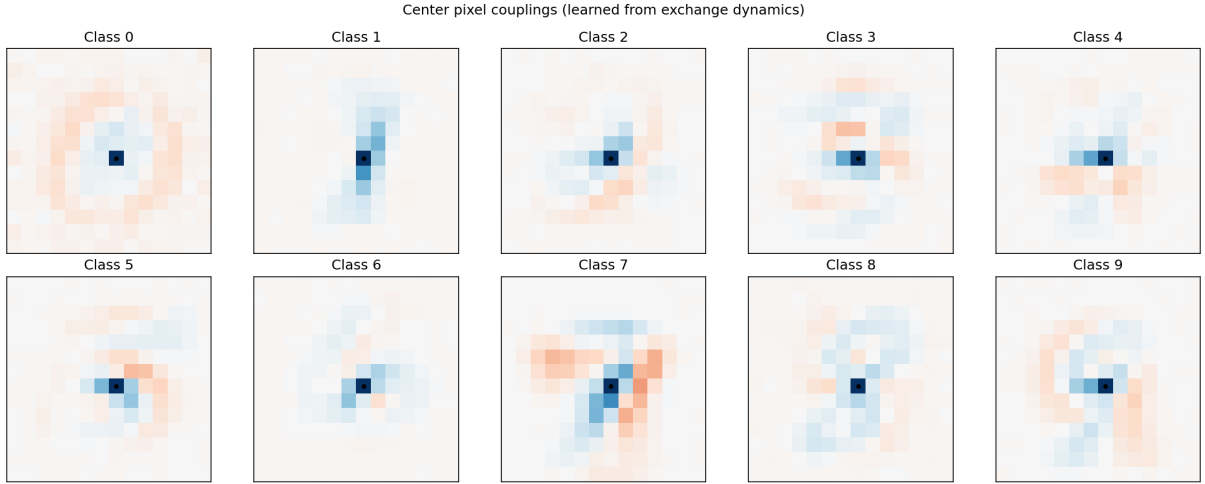

Figure 7: Learned couplings J_{ij} for each digit class, showing the coupling strength from the center pixel to all other pixels. Blue indicates positive coupling (correlated pixels), red indicates negative (anti-correlated). Unlike bias templates which resemble digit shapes, the coupling patterns show local correlation structure—confirming that the 78.8% accuracy on centered data comes from genuine second-order learning, not disguised template matching.
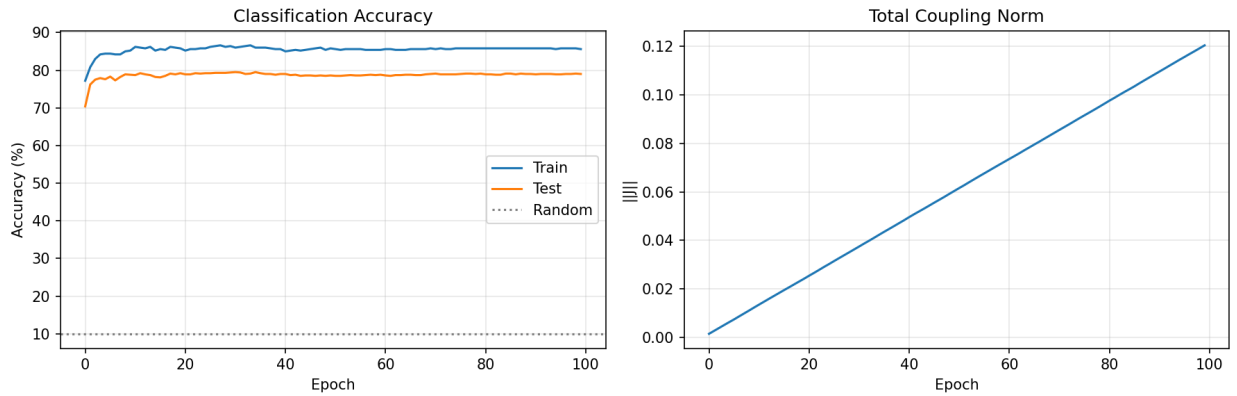

Figure 8: Learning curves for coupling-based classification. Left: Train and test accuracy over 100 epochs, showing rapid convergence to ∼78.8% test accuracy. The gap between train (∼86%) and test accuracy suggests room for regularization. Right: Total coupling norm $\|J\|$ grows linearly with training, indicating the optimizer continues to strengthen discriminative couplings.