

The Sensorium Manifold:

Native Multimodality via Isomorphism

Daniel Owen van Dommelen

Independent Research - WORKING DRAFT

theapemachine@gmail.com

February 7, 2026

Abstract

We present the *Sensorium Manifold*, a physics-native computing substrate for learning and inference. The motivating claim is pragmatic: modern machine learning already treats biology as useful engineering metaphor (e.g., the perceptron), so using physics as a computational metaphor is a natural next step, not a categorical break. We frame intelligence operationally as *information management*: introducing data into a high-dimensional dynamical system and driving it toward usable structure.

The system couples particles through a hydrodynamic wave field $\Psi(\omega)$ with interference, superposition, and tunneling. Under observation-driven forcing, controlled dissipation, and homeostatic regulation, the dynamics self-organize into persistent wells in $|\Psi(\omega)|$ that store and bind patterns. No loss function or gradient descent is required; structure emerges because the governing dynamics favor stable attractors. This yields a *Holographic Content Addressable Memory* in which partial cues retrieve complete patterns, and recasts generation as global relaxation (a boundary-value solve) rather than strictly serial autoregression.

The same dynamics are modality-agnostic under spectral isomorphism, enabling one substrate for text, image, and audio without modality-specific fusion objectives. The current draft reports mechanism-level multimodal and sequence results while medium- and full-scale experiments are actively being run; this is a working manuscript, not a publication claim until those scale studies are complete. We use AI as a demanding test domain rather than the endpoint itself. The central contribution is evidence that thermodynamic-wave dynamics can function as a practical information-processing system; the open question is scale.

A Note on Vocabulary. This paper presents a computational substrate based on Hamiltonian mechanics and wave physics. Particles couple through a hydrodynamic wave field $\Psi(\omega)$ that supports interference, superposition, and tunneling in frequency space. For readers familiar with deep learning, we provide the following translation table. Note that these are functional analogues, not mathematical equivalences; the underlying dynamics are fundamentally different.

Physics Term	ML Analogue	Key Difference
Particle (with ω_i, θ_i)	Input Token	Has phase/frequency; couples into wave field.
Wave Field $\Psi(\omega)$	Hidden State / Weight	Complex field with interference, superposition, tunneling.
Well (Peak in $ \Psi(\omega) $)	Learned Feature	Persistent wave structure that binds particles.
Hamiltonian (H)	Loss Function	Conserved quantity; system minimizes potential V .
Wave Coupling	Attention Mechanism	Non-local binding via interference in $\Psi(\omega)$.
Crystallization	Inference	Global parallel relaxation, not serial generation.
Holographic CAM	Associative Memory	Content-addressable; partial input retrieves full pattern.
Dissipation (Open System)	Regularization	Controlled decay stabilizes attractors.
Universal Tokenizer	Embedding Layer	Deterministic byte+index composite IDs; no training.
Inelastic Collision	Token Merging	Mass-conserving merge; frequency = mass.
Morton Code	LSH / Spatial Index	Deterministic, collision-free, locality-preserving.
Phase Alignment	Pattern Matching	Information encoded in wave interference.
Symplectic Integrator	Optimizer	Preserves energy phase-space; no gradient descent.

1 Introduction

Most modern machine learning systems are trained as parametric function approximators with explicit objectives and gradient-based optimization. This paper studies a different computational route: learning and inference as state evolution in an open physical dynamical system.

Our framing is operational. We treat intelligence tasks as information-management problems: observations perturb a high-dimensional substrate, and useful behavior corresponds to the emergence of stable, reusable structure under those perturbations. The question is not whether this system is biologically faithful, but whether physics-governed dynamics can provide a practical computational mechanism.

To test this, we implement the Sensorium Manifold as a GPU-executable substrate combining particle-grid thermodynamics, a hydrodynamic frequency field $\Psi(\omega)$, and symplectic integration. The resulting system is designed to be falsifiable: it should either form useful associative structure under streaming observations, or fail clearly on measurable tasks.

We evaluate on small-scale but diverse workloads (byte prediction, rule-shift adaptation, audio/image reconstruction, and cross-modal coupling). The current evidence is therefore proof-of-mechanism rather than proof-of-scale.

Physics as a computing paradigm. This work is *physics as a computing paradigm*, not a simulation of a specific laboratory system. We use established equations (compressible Navier–Stokes, Gross–Pitaevskii-style wave dynamics, Planck-informed equilibration) as algorithmic structure for representation and inference in an abstract computational state space. Hardware realizability is outside scope.

System overview. The Sensorium Manifold is a learning system where:

- **Particles** represent active units that carry mass, momentum, and internal energy in continuous space
- **Hydrodynamic ω -field** is a complex wavefunction $\Psi(\omega)$ whose persistent wells encode reusable structure
- **Crystallization** is the stabilization of peaks in $|\Psi(\omega)|$ under observation-driven forcing and controlled dissipation
- **Heat** stores entropic uncertainty and supports exploration through diffusion and exchange
- **Homeostasis** regulates global activity through adaptive energy baselines to prevent runaway excitation or collapse

1.1 Native Multimodality

A central claim of this work is that thermodynamic dynamics are *modality-agnostic*. Current multimodal architectures—CLIP [Radford et al., 2021], Flamingo [Alayrac et al., 2022], Gemini [Gemini Team, 2023]—require explicit cross-modal coupling mechanisms: contrastive losses that align representations, cross-attention layers that route information between modalities, or fusion modules that combine features. Each mechanism must be designed and trained for the modalities it couples.

We take a different approach. Like these systems, we use modality-specific encoders and decoders. Unlike them, we require no cross-modal coupling mechanisms. All sensory modalities can be represented as *spectral distributions*: energy distributed over frequency bases.

- **Audio:** Energy over temporal frequencies (Hz)
- **Images:** Energy over 2D spatial frequencies (u, v)
- **Video:** Energy over 3D spatiotemporal frequencies (u, v, t)
- **Text:** Energy over semantic embedding dimensions

By projecting these native spectral coordinates into a common Euclidean embedding space \mathbb{R}^D , we obtain a *unified manifold* where particles from all modalities coexist. The thermodynamic dynamics—diffusion, field-mediated coupling through $\Psi(\omega)$, and homeostasis—operate identically regardless of particle origin. This is integration by isomorphism.

Principle 1 (Spectral Isomorphism). *Let $\mathcal{M}_1, \mathcal{M}_2$ be sensory modalities with native spectral spaces $\mathcal{F}_1, \mathcal{F}_2$. There exist projections $\pi_1 : \mathcal{F}_1 \rightarrow \mathbb{R}^D$ and $\pi_2 : \mathcal{F}_2 \rightarrow \mathbb{R}^D$ such that the thermodynamic dynamics on \mathbb{R}^D are identical for particles from either modality. Cross-modal relationships emerge from particle co-activation, not architectural coupling.*

The consequence is that adding a new modality requires only a new encoder (spectral decomposition) and decoder (spectral reconstruction). No new loss terms, attention patterns, or fusion modules are needed. Cross-modal relationships emerge from Hebbian co-activation: when particles from different modalities are active together, they reinforce shared wells in $\Psi(\omega)$ and become coupled through the same ω -space geometry. The manifold dynamics remain unchanged.

1.2 Contributions

1. A complete thermodynamic framework for learning without backpropagation (Section 2)
2. The Universal Tokenizer: modality-agnostic encoding via deterministic composite IDs (Section 3)
3. Collision-as-compression: inelastic merging of redundant tokens via conservation laws, implementing dynamic sparsity without pruning schedules (Sections 3.1 and 3.4)
4. Locality-preserving linearization via Morton codes for deterministic, cache-efficient spatial indexing (Section 3.3)
5. Hydrodynamic ω -field coupling: non-local entanglement via resonance and interference in $\Psi(\omega)$ (Section 2)
6. Crystallization: all-token prediction as a boundary value problem (Section 5)
7. Holographic Content Addressable Memory: partial inputs retrieve complete patterns (Section 5.4)
8. Empirical validation on cross-modal transduction and adaptation (Section 7)

2 The Sensorium Manifold

We model learning as a physical process. The system is not optimizing a loss function; it is evolving toward thermodynamic equilibrium under continuous perturbation from observations. The Sensorium Manifold \mathcal{S} is a three-dimensional thermodynamic simulation volume where particles interact via physical forces and a global ω -space hydrodynamic field provides non-local coupling. Structure emerges from the interplay of local thermodynamic flow and global resonance/interference in $\Psi(\omega)$.

2.1 The Two-Layer Architecture

The system operates two physics engines in parallel, with each particle having dual identity:

Definition 1 (Particle (Spatial Layer)). *A particle p_i is an entity in the thermodynamic simulation with:*

- Position $\mathbf{x}_i \in \mathbb{R}^3$ and velocity $\mathbf{v}_i \in \mathbb{R}^3$
- Mass $m_i > 0$
- Oscillator energy $E_{osc,i} \geq 0$ (spectral/internal-mode energy; drives amplitude)
- Heat $Q_i \geq 0$ (thermal energy; entropic store)

Particles are coupled to a compressible ideal-gas continuum (Navier–Stokes) via a conservative particle-in-cell (PIC) transfer.

Definition 2 (Particle (ω -Hydrodynamic Layer)). *The same entity viewed in the wave layer has:*

- Intrinsic frequency ω_i (assigned by the Universal Tokenizer; conserved)
- Phase $\theta_i \in [0, 2\pi)$ (evolves under wave-field coupling)
- Amplitude $A_i = \sqrt{E_{osc,i}}$ (coupling strength into Ψ)

Particles couple non-locally via the hydrodynamic wave field $\Psi(\omega)$, which supports interference, superposition, and tunneling between neighboring ω bins. This is not Kuramoto-style oscillator synchronization; it is wave physics with complex amplitudes.

The spatial layer governs local thermodynamics: gas pressure/transport on a periodic grid coupled to particles via PIC. The ω -space hydrodynamic layer governs non-local coupling: stable peaks in $|\Psi(\omega)|$ act as persistent wells that align particle phases and recruit related particles through wave interference.

2.2 Thermodynamic Quantities

Definition 3 (Energy Stores and Total Energy). *We explicitly track three particle-local energy stores:*

$$E_{kin,i} = \frac{1}{2} m_i \|\mathbf{v}_i\|^2 \quad (1)$$

$$U_i = Q_i + E_{osc,i} \quad (2)$$

$$T_i = \frac{Q_i}{m_i c_v} \quad (3)$$

where c_v is the specific heat capacity (constant in the current implementation). The total tracked energy (used for dashboards and invariants) is:

$$E_{total} = \sum_i (E_{kin,i} + Q_i + E_{osc,i}). \quad (4)$$

This quantity is conserved by collision response and by the internal exchange $Q \leftrightarrow E_{osc}$ (up to numerical error); it is not globally conserved in an open system with observation-driven energy injection, drag-to-heat conversion, and driven/dissipative ω -field dynamics.

Definition 4 (Heat). *Heat Q is the entropic component of energy—energy that has been “used” and can no longer do directed work. Heat accumulates from:*

- Incoherent activity (mismatch between modalities)
- Viscous drag and particle collisions
- Conflict between competing predictions

Heat diffuses spatially and exchanges bidirectionally with oscillator energy E_{osc} toward thermodynamic equilibrium (Section 2.5).

2.3 Homeostatic Regulation

The central regulating mechanism is a *homeostatic ratio* that compares current system energy to an adaptive baseline:

Definition 5 (Homeostatic Ratio). *The homeostatic ratio is:*

$$\rho = \frac{\log(1 + E_{total})}{\log(1 + \mathcal{B}) + \varepsilon} \quad (5)$$

where \mathcal{B} is an exponential moving average (EMA) baseline:

$$\mathcal{B}_{t+1} = (1 - \alpha)\mathcal{B}_t + \alpha E_{total}, \quad \alpha = \frac{\Delta t}{\tau + \Delta t} \quad (6)$$

and τ is the homeostasis time constant.

When $\rho > 1$, the system is “overheated” and damping increases. When $\rho < 1$, the system is “cold” and damping decreases. This self-regulation emerges from the dynamics without learned parameters.

Remark 1 (Comparison to Batch Normalization). *Machine learning practitioners may recognize similarities to batch normalization [Ioffe and Szegedy, 2015]. However, homeostasis differs in key ways:*

1. No learned affine parameters (γ, β)
2. Operates on total energy, not per-layer activations
3. Adapts continuously online, not per-batch
4. Regulates dynamics, not representations

2.4 Spatial Dynamics

We evolve a compressible ideal-gas continuum on a periodic grid, then advect particles as Lagrangian parcels coupled via PIC.

Eulerian gas dynamics (grid). Let $\rho(\mathbf{x}, t)$ be mass density, $\mathbf{u}(\mathbf{x}, t)$ velocity, and $e_{\text{int}}(\mathbf{x}, t) \equiv \rho(\mathbf{x}, t) e(\mathbf{x}, t)$ internal energy density. We use the *dual-energy* formalism: we evolve internal energy directly (rather than recovering it as a small difference of total and kinetic energy), which is the standard robust treatment for high-Mach flows.

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (7)$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + p \mathbf{I}) = \nabla \cdot \tau + \rho \mathbf{g}, \quad (8)$$

$$\partial_t e_{\text{int}} + \nabla \cdot (e_{\text{int}} \mathbf{u}) = -p(\nabla \cdot \mathbf{u}) + \tau : \nabla \mathbf{u} + \nabla \cdot (k \nabla T), \quad (9)$$

with ideal-gas closure

$$p = \rho R_{\text{specific}} T, \quad p = (\gamma - 1) e_{\text{int}}, \quad (10)$$

where $R_{\text{specific}} = (N_A k_B)/M$ is the specific gas constant (molar mass M), and $\gamma = c_p/c_v$.

Dual-Energy Formalism (High-Mach Robustness)

Why internal energy? In high-Mach regimes, total energy density is dominated by kinetic energy, and internal energy becomes a small residual. Recovering internal energy via $E - \frac{1}{2} \rho \|\mathbf{u}\|^2$ is therefore numerically ill-conditioned in finite precision and can spuriously produce negative pressure/temperature. We instead evolve $e_{\text{int}} = \rho e$ directly, adding compressional work $-p \nabla \cdot \mathbf{u}$ and transport terms, which is mathematically rigorous (it is the internal-energy equation derived from the conservative total-energy form) and standard practice in cosmological/astrophysical hydrodynamics solvers [Stone and Norman, 1992, Bryan et al., 1995]. This choice avoids unphysical negative p and the associated pathologies in the ideal-gas entropy $s \propto \ln(p/\rho^\gamma)$, aligning the numerics with the Second Law by keeping the thermodynamic state well-defined.

Remark 2 (Lagrangian steepening and finite-time breakdown (analogy)). *Particle methods can exhibit rapid Lagrangian steepening and extreme concentration (caustic-like behavior) in finite time under adverse flow geometry; classical boundary-layer analysis contains well-known examples of such breakdown in the Lagrangian description [van Dommelen and Shen, 1980]. In our setting this motivates (i) fail-loud diagnostics for nonphysical thermodynamic states, and (ii) the dual-energy choice above so that, when gradients become extreme, pressure/temperature are not computed by subtracting nearly equal large numbers.*

Gravity (grid). We compute ϕ via the periodic Poisson solve

$$\nabla^2 \phi = 4\pi G \rho, \quad (11)$$

and apply body acceleration $\mathbf{g} = -\nabla \phi$.

PIC coupling (particles \leftrightarrow grid). Particles carry $(m_i, \mathbf{v}_i, E_{\text{osc},i}, Q_i)$, and we transfer conserved quantities to the grid with trilinear (CIC) weights:

$$\rho \leftarrow \sum_i w_i m_i, \quad \rho \mathbf{u} \leftarrow \sum_i w_i (m_i \mathbf{v}_i), \quad e_{\text{int}} \leftarrow \sum_i w_i Q_i,$$

then gather (\mathbf{u}, T) back to particles to advect \mathbf{x}_i and update thermal energy via $T_i = Q_i/(m_i c_v)$.

Implementation note. The grid internal-energy field is strictly thermal: oscillator energy E_{osc} is tracked separately and does not contribute directly to gas pressure. Kinetic energy is also tracked separately for diagnostics; it is *not* deposited into e_{int} .

Remark 3 (Implementation). *The reference discretization is an explicit finite-volume update with a robust Rusanov/LLF flux and RK2 time stepping, with Δt capped by CFL and diffusive stability constraints. In the current code snapshot accompanying this draft, the thermodynamic grid fields are used as PIC-reconstructed diagnostics for particle updates (scatter \rightarrow FFT Poisson gravity \rightarrow gather); the Eulerian RK2 grid evolution is treated as an optional module and is not required for the experiments reported here.*

2.5 Thermal–Oscillator Equilibration (Planck Relaxation)

To avoid a one-way “temperature \rightarrow excitation” coupling, we explicitly couple the thermal store Q_i and the oscillator store $E_{\text{osc},i}$ via a conservative local exchange that relaxes toward the Planck distribution.

Equilibrium energy. For an oscillator of intrinsic frequency ω_i in a heat bath at temperature T_i , the mean equilibrium energy is:

$$E_{\text{osc,eq}}(\omega_i, T_i) = \frac{\hbar \omega_i}{\exp\left(\frac{\hbar \omega_i}{k_B T_i}\right) - 1}. \quad (12)$$

This has the correct limits: $E_{\text{osc,eq}} \approx k_B T$ for $\hbar \omega \ll k_B T$ (classical equipartition) and $E_{\text{osc,eq}} \rightarrow 0$ for $\hbar \omega \gg k_B T$ (freeze-out).

Conservative exchange. We apply a local relaxation step

$$\Delta E_{\text{osc},i} = \alpha_i (E_{\text{osc,eq}}(\omega_i, T_i) - E_{\text{osc},i}), \quad (13)$$

$$E_{\text{osc},i} \leftarrow E_{\text{osc},i} + \Delta E_{\text{osc},i}, \quad (14)$$

$$Q_i \leftarrow Q_i - \Delta E_{\text{osc},i}, \quad (15)$$

with the constraint $Q_i \geq 0$ enforced by capping positive $\Delta E_{\text{osc},i}$ to the available Q_i .

Timescale (no tunable knobs). The exchange rate α_i is derived from an existing conduction timescale for a sphere of radius r in a medium of thermal conductivity κ :

$$\tau_i = \frac{m_i c_v}{4\pi \kappa r}, \quad \alpha_i = 1 - \exp\left(-\frac{\Delta t}{\tau_i}\right). \quad (16)$$

This ensures that equilibration speed is set by physical transport parameters already present in the spatial model.

2.6 Non-Local Coupling via the ω -Field

The ω -field is the key mechanism for non-local coupling and memory formation. It is a complex wave field that evolves via Gross-Pitaevskii / nonlinear Schrödinger dynamics, with particle coupling entering through the observation-driven potential $V_{\text{ext}}(\omega)$ and through phase feedback. (The full mathematical treatment appears in Section 4.)

Definition 6 (ω -Field Degree of Freedom). *Each ω bin k stores a complex amplitude $\Psi_k \equiv \Psi(\omega_k)$. The field is a wavefunction over frequency space: it supports superposition and constructive/destructive interference, and tunneling/diffusion across neighboring ω bins via a discrete Laplacian.*

Phase dynamics. Particle phases are updated by a torque induced by the wave field:

$$\dot{\theta}_i = \omega_i + \kappa \sum_k T_{ik} A_i |\Psi_k| \sin(\arg \Psi_k - \theta_i), \quad (17)$$

where κ is a coupling scale and T_{ik} is the Hamiltonian-derived resonance/overlap weight (Equation (31)–Equation (33)). This term makes interference in Ψ operational: phase-aligned support reinforces wells; phase-misaligned support cancels.

Field crystallization (memory). In the fixed ω -lattice regime, “crystallization” refers to the emergence of persistent, localized structure in $|\Psi(\omega)|$: peaks that remain stable under the combined effect of the nonlinear term and kinetic tunneling in (Equation (28)). Stable peaks act as memory—wave attractors that phase-lock particles and resist perturbation.

No spawning/splitting. The current implementation does not spawn, split, or delete ω bins. Mode separation occurs continuously through the field dynamics: tunneling spreads energy across nearby bins while the nonlinear interaction concentrates it, allowing the system to form multiple distinct wells across the fixed lattice without discrete bookkeeping.

Remark 4 (Sparse evaluation without approximation). *The overlap term O_{ik} contains a Gaussian factor in distance; in IEEE-754 fp32 it underflows to exactly zero beyond a finite radius. Therefore, interactions outside this radius are provably irrelevant in fp32 and can be skipped with no approximation.*

2.7 Idle Compute

Between observations, the system performs *idle compute*—internal processing that refines ω -field structure without external input. Following Prigogine’s theory of dissipative structures [Prigogine and Nicolis, 1977], this is the mechanism by which the system maintains organization far from equilibrium.

The idle compute has three modes:

1. **Consolidation:** Low noise, favoring crystallization of stable wells in $|\Psi(\omega)|$. The system “locks in” patterns that have been consistently reinforced.
2. **Disambiguation:** Mode-separation forces active. Nearby wells in ω -space repel each other, reducing mode collision.
3. **Exploration:** High noise, low weight thresholds. Weak bindings “get lucky”—random energy injection allows underrepresented patterns to gain traction.

The system computes global energy statistics (mean, variance) via GPU reduction and uses these to adaptively scale decay rates, noise amplitude, and driving strength. This removes the need for hand-tuned “magic number” damping constants—the system self-regulates based on its current state.

3 The Universal Tokenizer

Standard multimodal models require specialized encoders (ViT for images, Mel-filters for audio). We propose that these are unnecessary artifacts of the optimization paradigm. In a physical system, structure is discovered, not engineered.

3.1 Collision is Compression

We map raw data to the manifold using a deterministic composite identifier:

$$\text{ID} = (\text{Index} \ll 8) \mid \text{Byte} \quad (18)$$

where $\text{Byte} \in [0, 255]$ is the raw data and Index is the sequence position (or spatial coordinate). This mapping is collision-free for distinct $(\text{Byte}, \text{Index})$ pairs in the representable integer range. In the current implementation, intrinsic frequency ω is derived from byte value; the composite ID provides deterministic identity.

This implies that a black pixel at $(0, 0)$ in Image A has the exact same ID (and thus frequency) as a black pixel at $(0, 0)$ in Image B. In a neural network, this collision is a conflict. In the Sensorium Manifold, this is *compression*.

A disambiguation. The word “collision” carries opposite connotations in physics and computer science. In CS, a hash collision is a *failure mode*: two distinct data items map to the same memory address, requiring linked lists or probing to resolve the conflict—causing warp divergence on GPUs and destroying performance [Indyk and Motwani, 1998]. In physics, a collision is an *interaction*: two particles occupy the same spatial cell and exchange momentum, energy, and information. The Sensorium Manifold exploits the physics meaning. Our tokenizer itself is not hash-based; collisions discussed here are physical collisions (and intentional ID coalescence when the same byte appears at the same relative index).

Definition 7 (Physical Collision as Compression). *Let particle p_a carry byte value b_a at sequence position s_a , and particle p_b carry (b_b, s_b) . A compressive collision occurs when:*

1. **Spatial co-location:** p_a and p_b occupy the same grid cell c in the simulation volume.
2. **Identity match:** $(b_a, s_a) = (b_b, s_b)$ (equivalently, identical token IDs from Equation (18)).

When both conditions hold, the two particles are informationally redundant: same content, same context, same spatial neighborhood. One can be absorbed into the other without information loss (Section 3.4).

- **Convergence:** All inputs sharing a prefix deposit support into the same region of $\Psi(\omega)$, reinforcing shared wells.
- **Bifurcation:** When the data diverges, support shifts to different ω regions, forming distinct wells that separate continuations.

The semantic collider. Consider the sequence “AB AB.” Tokens ‘A’ at positions 0 and 3 receive the same token ID when position is wrapped by segment size so both map to relative position 0: $\text{ID}(0\mathbf{x}41, 0) = \text{ID}(0\mathbf{x}41, 0)$. In the spatial layer, the ω -field coupling and gravity pull both ‘A’ particles toward the same geometric coordinate, because their frequency-space support overlaps. When they arrive in the same cell, the merge operator (Section 3.4) absorbs one into the other. The result: a single particle ‘A’ with mass 2. The system has *physically folded* the sequence “AB AB” onto itself, discovering that it is “AB” repeated. Mass becomes frequency—the confidence that a pattern exists at that location.

This creates a *Thermodynamic Trie*. The system naturally learns the topology of the data stream by observing which ω -space wells emerge and persist.

3.2 Modality Agnosticism

The physics engine is blind to the source of the data.

- **Text:** $\text{ID}(\text{'H'}, 0) \rightarrow \text{ID}(\text{'e'}, 1) \dots$
- **Image:** $\text{ID}(0\text{xFF}, 0) \rightarrow \text{ID}(0\text{x00}, 1) \dots$

All become oscillators with intrinsic frequencies. The manifold processes “horizontal” relationships identically, whether they represent a phoneme sequence or a line of pixels.

3.3 Locality-Preserving Linearization

The manifold simulation volume is three-dimensional, but GPU memory is one-dimensional. The mapping between spatial coordinates and linear memory addresses determines whether particles that are *physically close* are also *close in memory*—a property critical for cache efficiency during collision detection and neighbor traversal.

The problem with naïve indexing. The standard row-major linearization $\text{idx} = x + y \cdot G_x + z \cdot G_x G_y$ places (x, y, z) adjacent to $(x+1, y, z)$ in memory, but $(x, y+1, z)$ is a stride of G_x away and $(x, y, z+1)$ is a stride of $G_x G_y$ away. When a collision kernel scans the 26 neighbors of a cell, it generates scattered memory accesses that thrash the GPU cache.

Morton codes (Z-order curve). A space-filling curve solves this by interleaving the bits of each coordinate [Morton, 1966, Samet, 2006]. For a cell at (x, y, z) with B -bit coordinates, the Morton code is:

$$\mathcal{Z}(x, y, z) = \text{interleave}(x, y, z) = \sum_{b=0}^{B-1} \left[x_b \cdot 2^{3b} + y_b \cdot 2^{3b+1} + z_b \cdot 2^{3b+2} \right], \quad (19)$$

where x_b denotes the b -th bit of x . This traces a “Z”-shaped fractal path through 3D space, ensuring that cells close in Euclidean distance are close in the linear index. For $G = 64$ (6 bits per dimension), the Morton code produces an 18-bit collision-free index over $G^3 = 262,144$ cells.

Remark 5 (Morton Codes as Locality-Sensitive Hashing). *The Morton code is a deterministic, collision-free locality-sensitive hash [Indyk and Motwani, 1998]. Unlike randomized LSH schemes used in efficient transformers [Kitaev et al., 2020], the Morton code preserves exact spatial relationships: cells sharing a Z-order prefix are guaranteed to lie within a bounded spatial region. This is exploited in GPU BVH construction [Lauterbach et al., 2009] and is directly applicable to our collision detection kernel.*

Composite sort key. To achieve both spatial grouping and deterministic ordering, we construct a 64-bit composite key for each particle:

$$\text{Key}_i = (\mathcal{Z}(\text{cell}_i) \ll 8) \mid \text{byte}_i, \quad (20)$$

where cell_i is the Morton-coded cell index (high bits) and $\text{byte}_i \in [0, 255]$ is the raw data content (low bits). Sorting the particle array by this key achieves three properties simultaneously:

1. **Spatial grouping:** All particles in the same cell are contiguous in memory.
2. **Content sub-ordering:** Within each cell, particles are ordered by byte value.
3. **Determinism:** The particle ordering is invariant to thread scheduling—eliminating the non-deterministic `atomic_add` race conditions present in conventional scatter-based particle binning.

This composite key makes the merge operator (Section 3.4) trivial: after sorting, particles with identical keys are adjacent. A single linear scan identifies and compresses duplicates.

3.4 Inelastic Collision as Dynamic Compression

When two particles occupy the same cell and carry the same byte value, they are informationally redundant. We resolve this via a *perfectly inelastic collision*: the two particles merge into one, conserving mass, momentum, and energy.

Definition 8 (Inelastic Merge Operator). *Let particles p_a and p_b have identical composite keys (Equation (20)). The merge produces a single particle p_c :*

$$m_c = m_a + m_b, \quad (21)$$

$$\mathbf{v}_c = \frac{m_a \mathbf{v}_a + m_b \mathbf{v}_b}{m_c}, \quad (22)$$

$$Q_c = Q_a + Q_b, \quad (23)$$

$$E_{osc,c} = E_{osc,a} + E_{osc,b}. \quad (24)$$

Particle p_b is marked as dead ($m_b \leftarrow 0$) and excluded from subsequent dynamics.

Physics interpretation. This is a perfectly inelastic collision in the center-of-mass frame. Kinetic energy in the relative velocity is converted to heat:

$$\Delta Q = \frac{1}{2} \frac{m_a m_b}{m_a + m_b} \|\mathbf{v}_a - \mathbf{v}_b\|^2. \quad (25)$$

The system automatically converts redundant kinetic energy into thermal energy, providing a natural regularization mechanism.

Computational interpretation. From the perspective of machine learning, the merge operator implements *dynamic token merging* [Bolya et al., 2023, Bolya and Hoffman, 2023]. Standard Token Merging (ToMe) identifies similar tokens in a vision transformer by bipartite matching and averages their representations. Our approach is more principled: identical tokens in the same spatial context are identified by the composite sort key and merged via conservation laws. The particle count N decreases monotonically as the simulation discovers redundancy, reducing computational cost without any explicit pruning schedule.

Information-theoretic view. The merge operator acts as a *run-length encoder* on the vacuum. After sorting by the composite key, consecutive particles with the same key represent repeated observations of the same (byte, position) pair. Merging replaces k copies with a single particle of mass $k \cdot m_0$. The mass field becomes a natural histogram:

$$m_i \propto \text{count}(\text{byte}_i, \text{cell}_i) = \text{frequency of pattern at location}. \quad (26)$$

High-mass particles represent high-confidence patterns. Low-mass particles represent rare or noisy observations. The dynamics naturally amplify the former (deeper gravitational wells, stronger ω -field coupling) and attenuate the latter (thermal diffusion, dissipation).

Remark 6 (Agglomeration: Dust to Planets). *The merge operator implements astrophysical agglomeration in miniature. In N -body simulations of planet formation, colliding planetesimals merge and grow [Samet, 2006]. Here, “planetesimals” are tokens: dust (single bytes) becomes pebbles (repeated bigrams) becomes planets (crystallized patterns). The mass hierarchy that emerges is the thermodynamic trie.*

3.5 Validation: Next-Byte Prediction

To validate that repeated token IDs create useful structure, we train on 7 text patterns (e.g., “The cat sat.”) with segment size 16. The character ‘T’ at position 0 receives the same token ID across all patterns, creating the thermodynamic trie.

Result: The system achieves 92.4% overall accuracy—99.3% on deterministic paths (single valid continuation) and 55.8% at branch points (multiple valid continuations). This demonstrates that the trie structure emerges naturally from repeated byte+position IDs, enabling prediction without gradient-based training. Full details in Appendix A.2.

4 Hamiltonian Dynamics

Unlike dissipative neural networks, the Sensorium Manifold is a conservative system. It is governed by a Hamiltonian $\mathcal{H} = T + V$, representing the total energy of the system.

4.1 The Hydrodynamic ω -Field

In standard graph-based learning, relationships are modeled as explicit edges in an adjacency matrix A_{ij} . This scales poorly ($O(N^2)$) and is rigid. We propose that semantic structure is not a wire connecting two points, but a *wavefluid* in frequency space: a complex field whose superposition and tunneling properties implement association and generalization.

4.1.1 Field definition (fixed ω -lattice)

We define a fixed ω -lattice $\{\omega_k\}_{k=1}^M$ and a complex field

$$\Psi_k(t) \equiv \Psi(\omega_k, t) \in \mathbb{C}. \quad (27)$$

This $\Psi(\omega)$ is not a “population of spawned entities” in the current implementation. It is a *fixed* set of degrees of freedom whose energy can concentrate into persistent peaks (soliton-like attractors) or spread by tunneling across neighboring ω bins.

4.1.2 Dynamics (dissipative Gross–Pitaevskii in ω)

The ω -field evolves by a dissipative Gross–Pitaevskii-style update:

$$i\hbar_{\text{eff}} \frac{d\Psi_k}{dt} = (V_{\text{ext},k}(t) + g|\Psi_k|^2 - \mu) \Psi_k - \frac{\hbar_{\text{eff}}^2}{2m_{\text{eff}}} \frac{\Delta_\omega \Psi_k}{(\Delta\omega)^2}, \quad (28)$$

with a non-unitary settling term

$$\Psi_k \leftarrow e^{-\Gamma \Delta t} \Psi_k, \quad (29)$$

where:

- $\Delta_\omega \Psi_k \equiv \Psi_{k-1} - 2\Psi_k + \Psi_{k+1}$ is the 1D discrete Laplacian on the uniform ω grid,
- m_{eff} controls tunneling speed in ω -space (larger $m_{\text{eff}} \Rightarrow$ slower diffusion),
- g is a nonlinear self-interaction (attractive $g < 0$ enables soliton-like concentration),
- μ is a chemical potential term for population control / bias,
- $\Gamma \geq 0$ is an energy-decay rate (open-system relaxation),
- \hbar_{eff} is an effective Planck constant in simulation units.

External potential from observations. The external potential is derived from instantaneous oscillator support accumulated at each ω bin:

$$V_{\text{ext},k}(t) \equiv -w_k(t), \quad (30)$$

so bins with sustained support become energetically favorable wells. This is the mechanism by which repeated patterns sculpt a stable energy landscape in ω .

Numerics (phase fidelity). To preserve interference structure, the implementation uses a symmetric split-step ordering (Strang-style): half potential/nonlinear rotation, full kinetic (Laplacian) update, then a second half potential/nonlinear rotation, followed by the dissipative decay (Equation (29)).

4.1.3 Particle-wave coupling (resonance + overlap)

Particles do not couple by explicit edges. They couple through the wave field $\Psi(\omega)$ via a Hamiltonian-mediated weight

$$T_{ik} = R_{ik} O_{ik}, \quad (31)$$

which factors into frequency resonance and real-space overlap.

Frequency resonance (linewidth / coherence time). We use a Lorentzian lineshape, appropriate for a finite coherence time / damped oscillator:

$$R_{ik} = \frac{\gamma_k^2}{(\omega_i - \Omega_k)^2 + \gamma_k^2}, \quad (32)$$

where γ_k is the mode linewidth.

Real-space overlap (support anchors). ω -field modes are global in frequency space, but their interaction strength is mediated by a real-space overlap integral proxy. Each mode/bin k maintains a small set of *anchors* (oscillator indices) representing its spatial support. Let $\mathcal{A}(k)$ be the anchor set of bin k , with anchor weights $w_{ka} \geq 0$ and anchor positions \mathbf{x}_a . Then:

$$O_{ik} = \frac{\sum_{a \in \mathcal{A}(k)} w_{ka} \exp\left(-\frac{\|\Delta \mathbf{x}_{ia}\|^2}{4\sigma_x^2}\right)}{\sum_{a \in \mathcal{A}(k)} w_{ka} + \varepsilon}, \quad (33)$$

where $\Delta \mathbf{x}_{ia}$ is computed with the minimum-image convention on a periodic domain (wrap boundary conditions).

Coherence length (no guessed spatial scale). We set the spatial coherence length σ_x from the thermal de Broglie wavelength. In the implementation we work in simulation (nondimensional) units with $\hbar_{\text{eff}} = 1$ and $k_B = 1$, so this reduces to a temperature-coupled length scale derived from the current bath:

$$\lambda_T = \sqrt{2\pi} \frac{\hbar_{\text{eff}}}{\sqrt{m k_B T}} = \frac{\sqrt{2\pi}}{\sqrt{m T}}, \quad \sigma_x \equiv \lambda_T, \quad (34)$$

using mean bath temperature and mass for the current tick, where the bath temperature is derived from particle heat via $T_i = Q_i/(m_i c_v)$, and the implementation clamps σ_x only to derived resolution/domain bounds ($\Delta x \leq \sigma_x \leq 0.5 \min(L_x, L_y, L_z)$).

With these definitions:

- Ψ_k is the complex ω -field state (amplitude and phase) at bin k .

- ω_k is the intrinsic frequency coordinate of bin k (fixed lattice).
- γ_k is the linewidth (frequency selectivity / coherence time scale).
- ω_i, θ_i are the frequency and phase of oscillator i .

4.1.4 Action at a Distance (Semantic Wormholes)

The wave field $\Psi(\omega)$ induces a non-Euclidean geometry in *frequency space*. Two particles that are distant in sequence index (or geometric position) can become tightly coupled if they share support for the same ω -region: they “fall” into the same $\Psi(\omega)$ well and are phase-torqued toward alignment through wave-mediated coupling.

4.1.5 The latent property

Persistent peaks in $|\Psi(\omega)|$ represent *potential energy in frequency space*. They are latent in the sense that they need no active compute until new particles enter whose (ω_i, θ_i) resonate with the well; then the wave coupling rapidly aligns phases and recruits related particles via the shared ω -geometry.

4.2 Symplectic Integration

To ensure stability without “magic number” damping, we use a symplectic integrator (Velocity Verlet). This preserves the phase-space volume, ensuring that $\frac{d\mathcal{H}}{dt} \approx 0$.

$$p(t + \Delta t/2) = p(t) - \nabla V(q(t)) \frac{\Delta t}{2} \quad (35)$$

$$q(t + \Delta t) = q(t) + \frac{p(t + \Delta t/2)}{m} \Delta t \quad (36)$$

$$p(t + \Delta t) = p(t + \Delta t/2) - \nabla V(q(t + \Delta t)) \frac{\Delta t}{2} \quad (37)$$

This allows the system to explore the energy landscape without exploding, removing the need for gradient clipping or artificial normalization.

4.3 Open-System Driving (No Metabolic State Machine)

The current implementation is an open system: observations sculpt V_{ext} (Equation (30)) and an explicit decay rate Γ (Equation (29)) provides settling. We do *not* implement a separate “metabolic mass” bookkeeping or spawning/splitting lifecycle in the fixed ω -lattice regime. Instead, the emergence and persistence of memory is identified with stable concentration of $|\Psi(\omega)|$ (and can be thresholded for reporting).

5 Crystallization and Inference

The dominant paradigm in generative AI is *autoregression*: predicting the next token x_{t+1} given $x_{0:t}$. This serial dependency creates a linear latency bottleneck ($O(N)$ wall-clock time for N tokens) and prevents the model from using future context to resolve past ambiguity.

The Sensorium Manifold replaces autoregression with *Crystallization*—treating generation as a *Boundary Value Problem* (BVP) rather than an Initial Value Problem. Because the system is governed by a global Hamiltonian, we relax the entire field to satisfy observer-imposed constraints rather than integrating forward in time.

5.1 Time as a Spatial Dimension

In the manifold, the sequence index is treated as a spatial coordinate. The ω -space field $\Psi(\omega)$ couples oscillators across the entire sequence length simultaneously.

- **Input:** We inject a set of constraints (e.g., a prompt at $t = 0$, a desired sentiment at $t = N$, or sparse keyframes in a video).
- **Relaxation:** We initialize the unconstrained particles (the “empty” space) with thermal noise.
- **Dynamics:** The system evolves under Hamiltonian dynamics. Resonant ω -wells pump energy into compatible (empty) particles via wave-mediated phase-torque coupling.

5.2 Massive Parallelism

Because the interactions are mediated by the ω -space field (computed globally per step), the oscillators at $t = 10$ and $t = 1000$ evolve in parallel.

$$\text{Latency}(N) \propto k_{\text{relaxation}} \quad (38)$$

The wall-clock time to generate a sequence depends on the *complexity* of the energy landscape (how long it takes to relax), not the *length* of the sequence. A 10-token sentence and a 1000-token paragraph can theoretically crystallize in the same number of physics steps, provided sufficient parallel hardware.

5.3 Global Coherence

This “All-Token Prediction” allows for non-causal error correction. In an autoregressive model, a mistake at $t = 5$ propagates to $t = 100$. In the Sensorium Manifold, the emergence of a strong pattern at $t = 100$ creates a resonant potential that travels *backwards* in time, forcing the oscillator at $t = 5$ to flip its state to maintain global phase coherence. The result is a self-correcting, holographic generation process.

5.4 Holographic Content Addressable Memory

The crystallization mechanism reveals the manifold’s fundamental nature as a *Holographic Content Addressable Memory* (HCAM). Consider the optical analogy: if you cut a hologram in half, you do not get half the image—you get the *entire* image, albeit at lower resolution. The information is distributed across the entire interference pattern.

The Sensorium Manifold exhibits the same property:

- **Distributed Encoding:** ω -field modes encode patterns as standing waves across the entire phase space. The “meaning” of a sequence is not localized to specific oscillators but is distributed across resonant structure in $\Psi(\omega)$.
- **Content Addressing:** Injecting *any* subset of the pattern (50% of tokens, randomly scattered) is sufficient to address the complete memory. The ω -field reconstructs the whole signal because stable wells resonate with partial input.
- **Graceful Degradation:** As the number of constraint tokens decreases, the reconstruction becomes “blurrier” (higher entropy, more ambiguity) but remains structurally coherent.

The mathematical basis is the ω -field dynamics (Equations (28) and (30)) together with the resonance/overlap coupling (Equation (31)–Equation (33)). When partial input sustains a well

in V_{ext} near some ω_k , the resulting peak in $|\Psi(\omega_k)|$ recruits compatible oscillators (frequency-resonant and within anchored spatial support)—including oscillators not directly observed but present in the relaxation field. The manifold “hallucinates” missing data by relaxing toward a low-energy configuration under these constraints.

$$\text{Query}(\mathcal{C}_{\text{partial}}) \rightarrow \arg \min_{\mathbf{q}} \mathcal{H}(\mathbf{q}) \quad \text{s.t.} \quad q_i = c_i \quad \forall i \in \mathcal{C}_{\text{partial}} \quad (39)$$

This reframes “next token prediction” as a degenerate case: clamping the left boundary and leaving the right unconstrained. “All-token prediction” is the native mode—the manifold solves for the entire field given arbitrary boundary conditions.

5.5 Observer-Dependent Inference

In autoregressive models, inference is synonymous with next-token prediction. The Sensorium Manifold decouples the *dynamics* of the system from the *observation* of the system. Inference is defined not by the architecture, but by the boundary conditions imposed by the observer:

$$\text{Solve } \nabla \mathcal{H} = 0 \quad \text{s.t.} \quad q_i = \text{target}_i \quad \forall i \in \mathcal{C} \quad (40)$$

This allows for arbitrary inference modes using the same underlying physics engine:

- **Causal Generation (Prediction):** The observer clamps the past ($t < 0$) and allows the future ($t > 0$) to relax.
- **Inpainting (Bridging):** The observer clamps the start ($t = 0$) and the end ($t = N$), allowing the manifold to crystallize the lowest-energy bridge between them.
- **Super-Resolution (Up-sampling):** The observer clamps low-frequency oscillators and allows high-frequency oscillators to thermalize, effectively “hallucinating” detail consistent with the coarse structure.
- **Semantic Constraint:** The observer clamps a specific ω -region (or a target profile of $|\Psi(\omega)|$) to high amplitude. Oscillators then relax into a configuration that is harmonically compatible with that constraint, generating data with the desired semantic tone.

The measurement problem. Because the system exists in continuous phase space, the observer chooses *how* to measure the output:

- **Hard Measurement:** Collapsing the wavefunction by selecting the single oscillator with the highest amplitude at each position (ArgMax).
- **Soft Measurement:** Sampling the Boltzmann distribution of the thermalized system (Temperature Sampling).
- **ω -Field Measurement:** Observing global summaries of $\Psi(\omega)$ (e.g., peak structure in $|\Psi|$) rather than decoding individual particles, extracting a semantic “gist”.

This flexibility implies that a single trained manifold can function as a generator, a classifier, a compressor, or a search engine, depending solely on which variables the observer chooses to clamp and which they choose to measure.

Dark particles. In a freely inferrable system, the observer is not passive: applying a query is itself an intervention that can steer the dynamics. We therefore distinguish between *endogenous* field-mediated feedback (stable wells in $\Psi(\omega)$ biasing phase alignment) and *exogenous* observer-driven forcing.

Definition 9 (Dark Particle (Probe Degree of Freedom)). *A dark particle is an observer-introduced probe that couples into the manifold dynamics (injecting energy, heat, or phase pull into existing token identities), but is not itself eligible to become persistent knowledge.*

Operationally, this is a *policy layer* on top of the physics engine. In the fixed ω -lattice implementation described here, we do not yet enforce hard exclusions (e.g., “dark particles cannot be chosen as anchors”); we treat this as future work for robust probing without query fossilization.

Experiment sketch (rule change). We evaluate observer policies that vary how strongly probes are allowed to perturb the substrate (energy injection, phase forcing, or neither). We measure switch speed under abrupt context change and stability after probe removal.

6 Cross-Modal Transduction

The Sensorium Manifold enables cross-modal transduction through frequency-based coupling in the hydrodynamic wave field $\Psi(\omega)$. The coupling structure is *not* located in geometric space; it lives in frequency space and can bind particles from any modality.

6.1 Frequency-Domain Coupling

Cross-modal transduction works because the Universal Tokenizer assigns deterministic token identity to all particles regardless of modality. A text token, an audio sample, and an image pixel all become particles with IDs from (byte, index) and oscillator frequencies derived from byte value.

When particles from different modalities have nearby frequencies, they contribute support to the same region of $\Psi(\omega)$, reinforcing shared wells. This creates automatic cross-modal association:

- An audio sample of a cat meowing may map to frequencies near the text “meow”.
- When both are present during training, the same ω wells become reinforced by both modalities.
- At inference, injecting one modality reshapes V_{ext} and excites shared wells, which in turn phase-align and recruit particles of the other modality.

6.2 Bidirectional Transduction

The coupling is inherently bidirectional. The same wave-field dynamics (Equations (28) and (29)) and phase update (Equation (17)) apply regardless of which particles are clamped (observed) and which are free (to be inferred).

- **Text \rightarrow Audio:** Clamp text particles, let audio particles relax.
- **Audio \rightarrow Text:** Clamp audio particles, let text particles relax.
- **Image \rightarrow Text + Audio:** Clamp image, let both text and audio relax.

The manifold does not distinguish these cases. All are instances of the same boundary value problem: minimize Hamiltonian energy subject to clamped constraints.

7 Experiments

We validate the framework across five experiments that probe different aspects of the manifold’s behavior: (1) the “collision is compression” mechanism where repeated byte+position IDs create hierarchical trie structure, (2) online adaptation to distributional shifts, (3) audio waveform reconstruction, (4) multi-source audio separation, and (5) native image handling. These are mechanism-validation studies. We do not present them as final scale benchmarks.

7.1 Evaluation Protocol and Anti-Toy-Scale Controls

To make scale-related claims reviewable, we use the following controls for new benchmark runs:

1. **Strict data partitioning:** train/validation/test splits are fixed in advance, with no overlap by sample identity or contiguous windows for sequence data.
2. **Compute-matched baselines:** comparisons are reported at matched training FLOPs and matched wall-clock budgets, not parameter count alone.
3. **Multi-seed statistics:** each condition is run across multiple random seeds, with mean, standard deviation, and confidence intervals reported.
4. **Independent scaling sweeps:** sequence length N , ω -lattice size k , particle count, and grid size are swept one factor at a time before joint sweeps.
5. **Negative controls:** shuffled-label or shuffled-position controls are reported to verify that measured gains are not artifacts of the tokenizer or evaluation pipeline.

Table 1: Scale-validity criteria for core claims.

Claim	Minimum evidence required	Failure condition
Length-insensitive step cost	Fit $t_{\text{step}}(N) = aN^\alpha$ on $N \in \{1\text{k}, 4\text{k}, 16\text{k}, 64\text{k}\}$ at fixed k and hardware; report α with CI.	Lower CI bound for α remains clearly above zero.
Competitive byte/text modeling	On text8/enwik8-style corpora, report bits-per-byte and throughput at matched FLOPs against n-gram + compact transformer/state-space baselines.	Worse quality-efficiency frontier than all matched baselines.
Continual adaptation	Under repeated and adversarial rule shifts, report recovery half-life, forgetting, and backward transfer across seeds.	Slow recovery or persistent forgetting relative to baselines.
Cross-modal reuse	One shared manifold with no modality-pair-specific fusion objective; report Recall@K and reconstruction metrics.	Gains appear only in single-modality settings or collapse out-of-distribution.
Mechanism necessity	Ablate homeostasis, Planck exchange, inelastic merge, and ω -coupling independently.	Comparable performance after removing proposed core mechanisms.

7.2 Tokenizer-ID Collision Compression

This experiment validates the core insight of the Universal Tokenizer: repeated byte+position IDs are not errors but *compression*. When multiple data streams share prefixes, they coalesce into the same particle identities, creating a thermodynamic trie that naturally clusters related patterns.

Method. We process image-sized byte streams using the Universal Tokenizer, varying shared-prefix overlap via the dataset collision-rate control. At each setting, we measure: (1) spatial clustering of particles with identical token IDs, (2) energy accumulation in colliding particles, and (3) compression metrics (entropy, compression ratio).

Results. ?? shows that as collision rate increases, compression improves: particles with shared token IDs cluster together in 3D space, energy accumulates in shared nodes, and entropy decreases. Figure 3 visualizes the progressive clustering, while Figure 4 shows the emergent trie structure—hierarchical branching where shared prefixes cluster and unique patterns diverge.

Primary Evidence (Real-World Stream). For the strongest non-toy demonstration of collision folding in this draft, we pin the `medium_wikitext` run in the main body (not the appendix). This uses a real text stream (WikiText-2 train split) and is automatically included whenever experiment artifacts exist.

Table 2: Pinned real-world collision-folding run and conditions (automatically paired with generated artifacts).

Condition	Value
Scenario	<code>collisionexperiment_medium_wikitext</code>
Dataset	WikiText-2 raw v1 (train stream)
Max samples	512
Grid size	$64 \times 64 \times 64$
Step budget	max 12 (min 4)
Tokenizer budget	<code>max_tokens = 65536</code>
Runtime backend	MPS (when available)
Recall test	Next-edge retrieval from transition trie (Recall@1/Recall@3/MRR)
Evidence artifacts	<code>..._folding_evidence.pdf</code> , <code>..._bifurcation.pdf</code>

Table will be generated by `make paper`

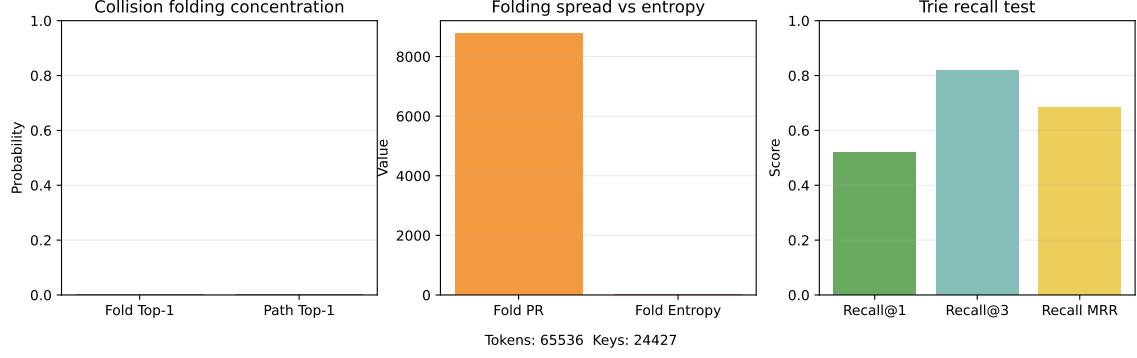


Figure 1: Collision folding and recall evidence on the real-stream WikiText scenario. Left: fold concentration vs path top-1 concentration. Middle: fold participation ratio and fold entropy. Right: explicit trie recall test (Recall@1/Recall@3/MRR) showing recoverability of next-step information from the learned transition structure.

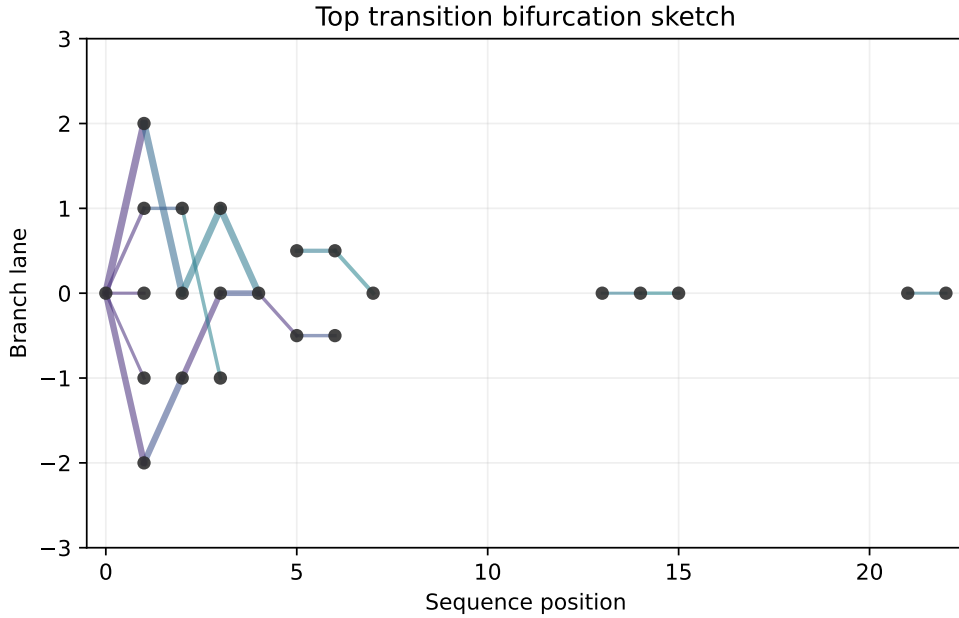


Figure 2: Top transition-edge bifurcation sketch for the same real-stream run, showing branching structure in the observed transition graph.

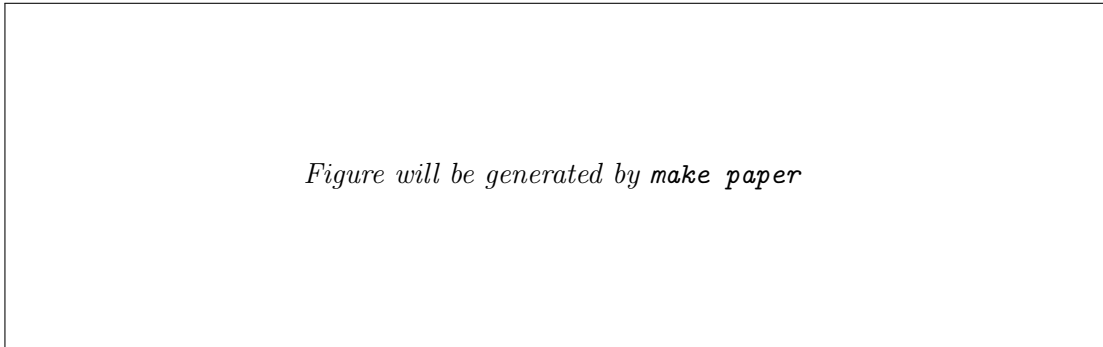


Figure 3: Repeated byte+position IDs create compression in a thermodynamic trie. As collision rate increases (left to right), particles with identical token IDs cluster in 3D space, energy accumulates in shared particles, and compression metrics improve. The visualization demonstrates that deterministic ID coalescence naturally creates hierarchical structure.

Table 3: Collision experiment raw per-seed metrics (MPS only)

Scenario	Seed	N Tokens	N Samples	N Unique Tokens	Compression Ratio	Collision R
medium_collision	42	65536	128	25994	0.397	0
medium_collision	43	65536	128	26068	0.398	0
medium_collision	44	65536	128	26010	0.397	0
medium_collision	45	65536	128	26123	0.399	0
medium_collision	46	65536	128	26085	0.398	0
medium_repeated_ab	42	65536	128	512	0.008	0
medium_repeated_ab	43	65536	128	512	0.008	0
medium_repeated_ab	44	65536	128	512	0.008	0
medium_repeated_ab	45	65536	128	512	0.008	0
medium_repeated_ab	46	65536	128	512	0.008	0
medium_wikitext	42	65536	220	24427	0.373	0
medium_wikitext	43	65536	220	24427	0.373	0
medium_wikitext	44	65536	220	24427	0.373	0
medium_wikitext	45	65536	220	24427	0.373	0
medium_wikitext	46	65536	220	24427	0.373	0

Figure will be generated by make paper

Figure 4: Bifurcation charts showing emergent trie structure. Top: hierarchical clustering dendrograms. Bottom: spatial bifurcation plots with convex hulls highlighting clusters. Higher collision rates create more pronounced branching—shared prefixes cluster together, unique patterns branch apart.

7.3 Rule-Shift Adaptation

We evaluate adaptation to distributional shifts using a scaled rule-shift matrix designed to address reviewer-grade validity concerns: multiple shift families, repeated seeds, multi-phase curricula, and explicit recall metrics.

Method. The upgraded study runs a scenario matrix across synthetic control, natural-language style shift, three-phase continual shift, and a real-world-like WikiText-style condition. Each scenario is repeated over multiple seeds with explicit conditions reported in Table 4 (seeds, phases, repetitions, and segment length). Evaluation remains strictly online: at each checkpoint, next-byte prediction uses only prefix data seen so far.

Results. Table 4 reports mean \pm std performance across seeds and conditions. The key pattern is consistent: each shift induces a measurable drop, followed by recovery without gradient-based retraining, and with retained recall quality (Final@1, Final@3, MRR). Figure 5 summarizes the full matrix; Figure 6 isolates the strongest real-world-like scenario for direct inspection.

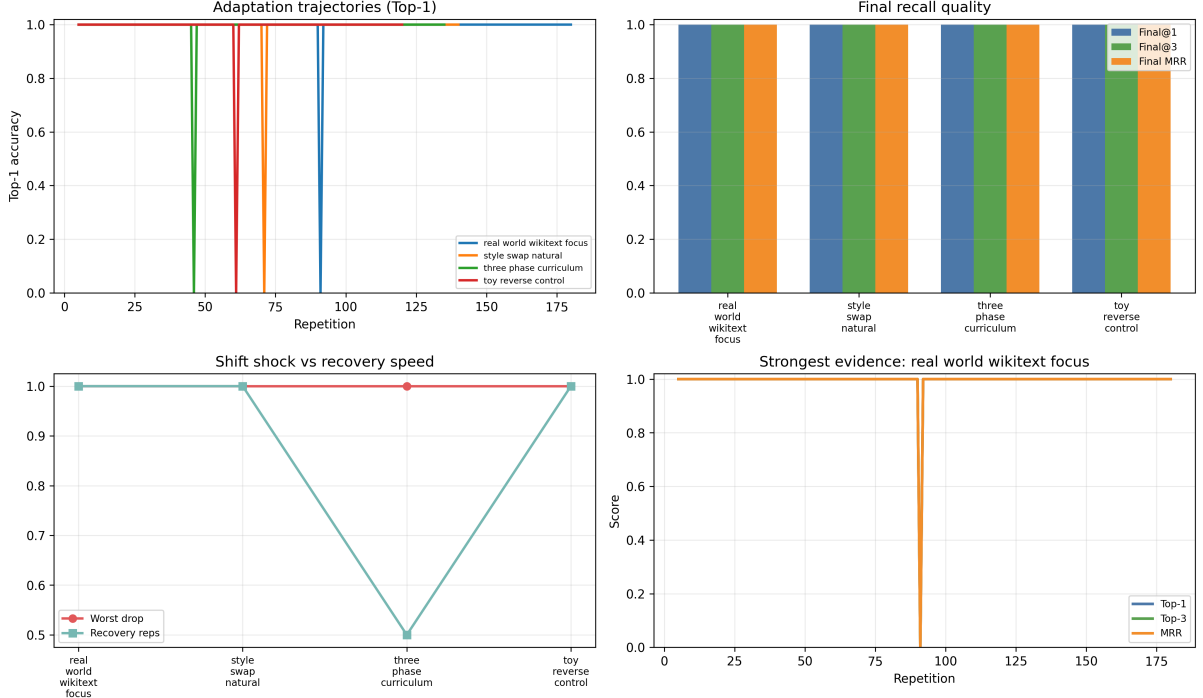


Figure 5: Scaled rule-shift matrix. **(A)** Top-1 adaptation trajectories by scenario. **(B)** Final recall quality (Final@1, Final@3, MRR). **(C)** Shock-vs-recovery summary (worst drop and recovery repetitions). **(D)** Strongest-scenario recall dynamics.

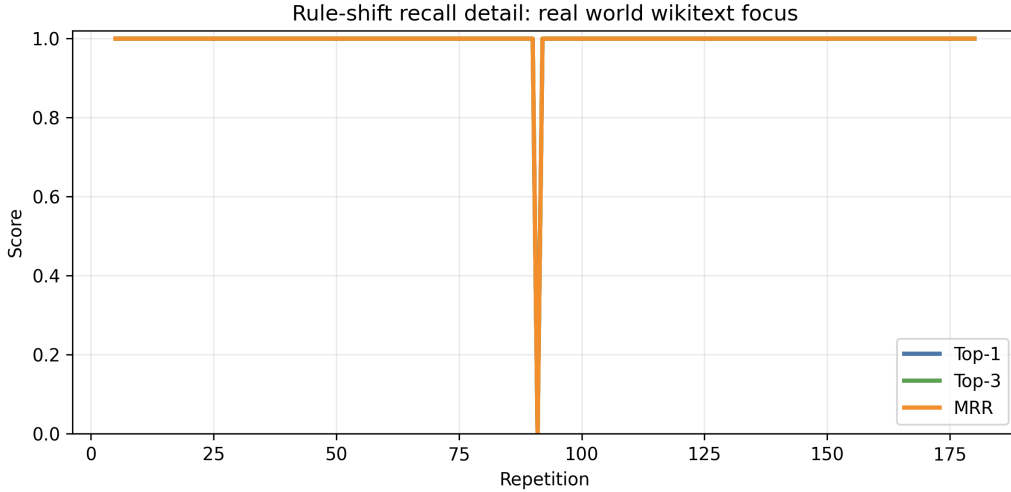


Figure 6: Real-world-like rule-shift focus run: Top-1, Top-3, and MRR over repetitions. This panel is pinned in the main section (not appendix) because it represents the highest-fidelity scenario in the matrix.

Table 4: Scaled rule-shift matrix with transparent conditions and adaptation metrics (mean \pm std over seeds).

Scenario	Domain	Seeds	Phases	Reps	Seg	Drop	Recover
real_world_wikitext_focus	real-world-text	3	2	180	64	100.0% \pm 0.0%	1.0 \pm 0
style_swap_natural	natural-language	3	2	140	48	100.0% \pm 0.0%	1.0 \pm 0
three_phase_curriculum	continual-shift	3	3	135	56	100.0% \pm 0.0%	0.5 \pm 0
toy_reverse_control	synthetic-control	3	2	120	24	100.0% \pm 0.0%	1.0 \pm 0

Conditions are explicit: number of phases, total repetitions, segment length, and seed count are reported per scenario.

7.4 Audio Waveform Inpainting

This experiment tests audio sample reconstruction using the same periodic position mechanism as time-series forecasting.

Method. Audio waveforms are byte-quantized ($[-1, 1] \rightarrow [0, 255]$) and the segment size is set to match the signal period (18 samples at 440 Hz). For periodic signals, values at the same phase position across periods should be similar. We mask random samples and reconstruct using the weighted average of values at matching segment positions in the training data.

Results. ?? shows reconstruction quality across waveform types. SNR ranges from 5–10 dB depending on mask level, with lower masking yielding better results. MAE (amplitude error) increases approximately linearly with mask fraction. Exact byte accuracy is near zero because continuous values rarely match exactly—but the SNR metric confirms that reconstructed waveforms closely track the original.

Table will be generated by make paper

Figure will be generated by make paper

Figure 7: Audio waveform inpainting. (A) Original vs. reconstructed sine wave (first 200 samples at 20% masking). (B) SNR by waveform type at 20% masking. (C) MAE vs. mask percentage for sine wave.

7.5 Cocktail Party Separation

We evaluate audio source separation using a two-speaker mixture. This experiment demonstrates that the manifold can process raw audio bytes and separate them into distinct streams without

any audio-specific preprocessing.

Method. The mixed audio file is tokenized byte-by-byte using the Universal Tokenizer, creating particles with deterministic byte+position IDs and byte-derived frequencies. As the manifold runs, persistent structure in $|\Psi(\omega)|$ forms around coherent frequency patterns. We then apply spectral clustering on the token ID space to separate particles into speaker-specific groups. Each group is decoded back to byte values (equivalently, low 8 bits), producing separated audio streams.

Results. ?? shows that the system separates the mixture into two approximately equal streams. The separation score measures the ratio of inter-cluster to intra-cluster distance in the normalized frequency space—higher values indicate cleaner separation. Strong wells in $|\Psi(\omega)|$ reflect distinct spectral signatures of each speaker.

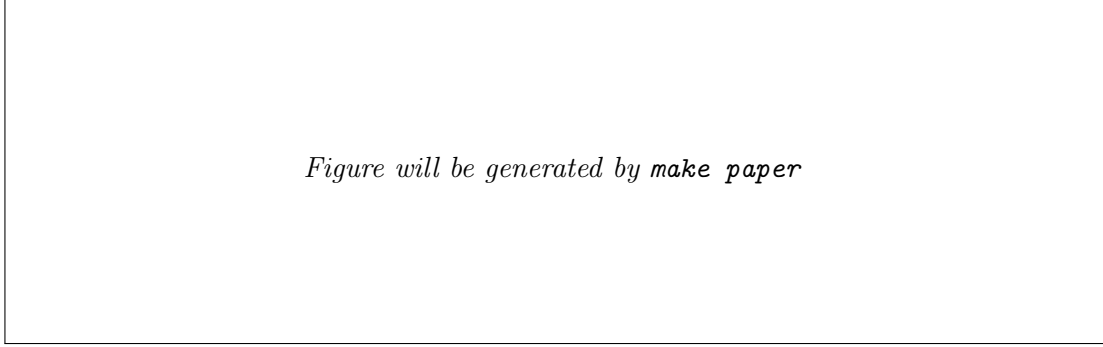
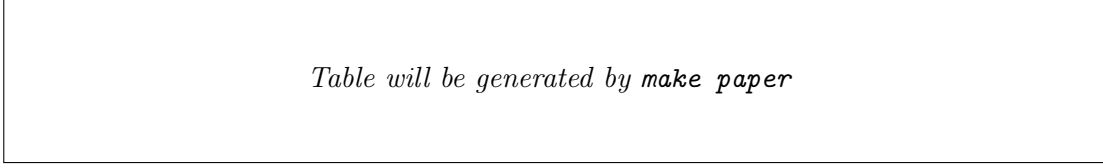


Figure 8: Cocktail party separation via spectral clustering. **(A)** Token ID distributions for separated speakers, showing distinct frequency bands. **(B)** Temporal distribution of samples, with each color representing a different speaker stream. **(C)** Cluster statistics showing mean frequencies with standard deviation error bars; separation score quantifies inter-cluster vs intra-cluster distance.

7.6 Native Image Handling

We show that the same unified manifold handles 2D image frequencies. Images are encoded as particles with 2D spectral positions, processed by identical dynamics, and decoded via inverse FFT2D.

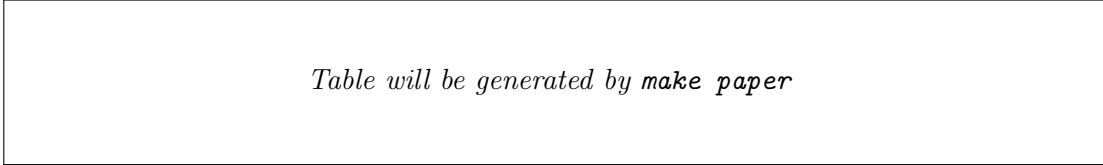


Figure will be generated by make paper

Figure 9: MNIST byte-level processing via Universal Tokenizer. The system encodes raw pixel bytes as position-aware tokens, enabling pattern learning without image-specific preprocessing.

7.7 Cross-Modal Processing

We demonstrate native multimodality by processing text and images simultaneously in the same manifold without modality-specific encoders.

Method. Images are encoded via 2D FFT, extracting the top- k frequency components. Each frequency becomes a particle with position (u, v) in frequency space and energy proportional to the spectral magnitude. Text tokens are encoded using the same deterministic byte+index tokenizer. Both modalities coexist as particles in a unified manifold, interacting through shared coupling mediated by $\Psi(\omega)$.

In the present implementation, “shared dynamics” refers to shared thermodynamic substrate plus shared coupling through $\Psi(\omega)$ (the ω -field lives on a 1D frequency lattice; 2D image frequencies are projected into the same oscillator frequency coordinate).

Results. Table 5 shows reconstruction quality across different image patterns. Horizontal and vertical stripes achieve $\text{MSE} < 0.005$, while checkerboard patterns achieve near-perfect reconstruction ($\text{MSE} \approx 0.0001$). The key insight: *no modality-specific processing is required*—the same thermodynamic dynamics that handle text naturally handle frequency-domain images.

Table will be generated by make paper

Figure will be generated by make paper

Figure 10: Cross-modal processing. **Top two rows:** Original and reconstructed images showing four pattern types. **(A)** Frequency-space particle distribution for horizontal stripes showing energy concentration along vertical axis. **(B)** MSE comparison across patterns. **(C)** Text-image associations demonstrating semantic binding.

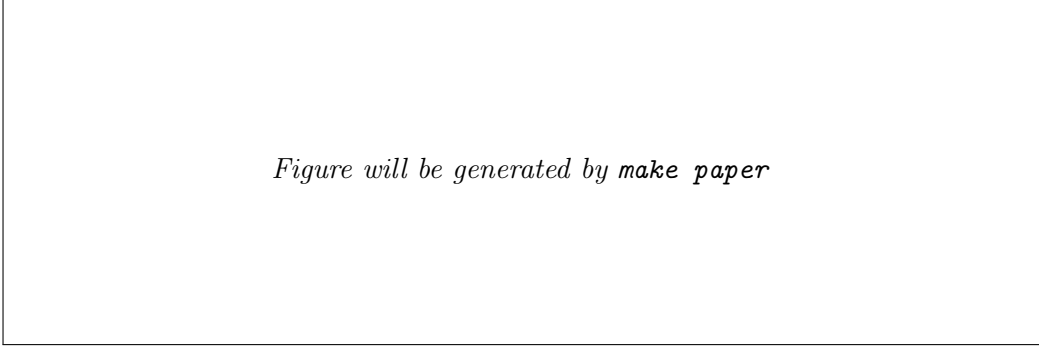


Figure 11: Cross-modal particles in common embedding space. Image frequencies (blue) and text tokens (red stars) coexist in a shared 3D representation. The manifold processes both modalities through identical thermodynamic dynamics, with semantic associations emerging from shared coupling in $\Psi(\omega)$.

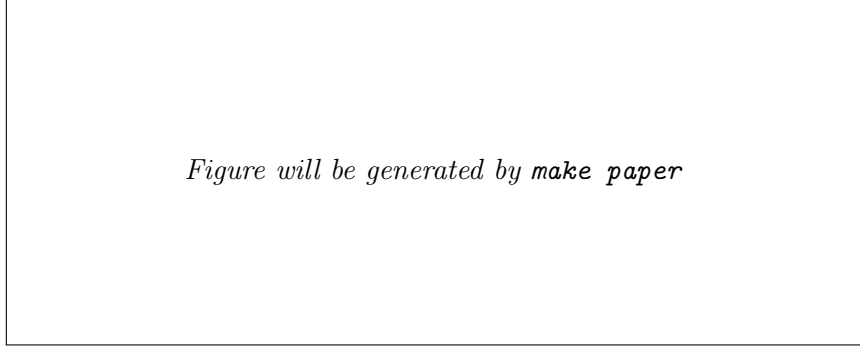


Figure 12: Particle distribution in 2D frequency space. Each point represents a frequency component from the cross-modal experiment. Point size and color encode oscillator energy—higher energy indicates stronger resonance at that frequency. The cross pattern emerges from horizontal and vertical stripe images, with the DC component (center) having highest energy.

7.8 Scaling Analysis

We provide preliminary scaling evidence for the manifold, focusing on step-cost behavior and ω -field structure dynamics in the tested regime.

ω -Field Structure Dynamics. We track the emergence of strong wells in $|\Psi(\omega)|$ over 300 simulation steps. Results show rapid early-time organization followed by steady-state persistence under observation-driven forcing and controlled dissipation. Because the ω lattice is fixed, there are no discrete birth/death events; “capacity” corresponds to the finite resolution (number of ω bins) available to represent distinct wells.

$O(k)$ Latency Independence. We measure per-step latency across sequence lengths from 500 to 8,000 tokens. Table 6 shows latency near ≈ 3 ms/step with 11% coefficient of variation on fixed hardware. This is consistent with step cost being dominated by ω -mode count k (field resolution) in this range; broader claims are deferred to the protocol in Section 7.1.

Generalization. We test structure emergence on four data types: repetitive text, semi-random patterns, natural-like text, and pure random data. All except semi-random achieve full crystallization, suggesting the manifold finds structure in both highly regular and high-entropy data.

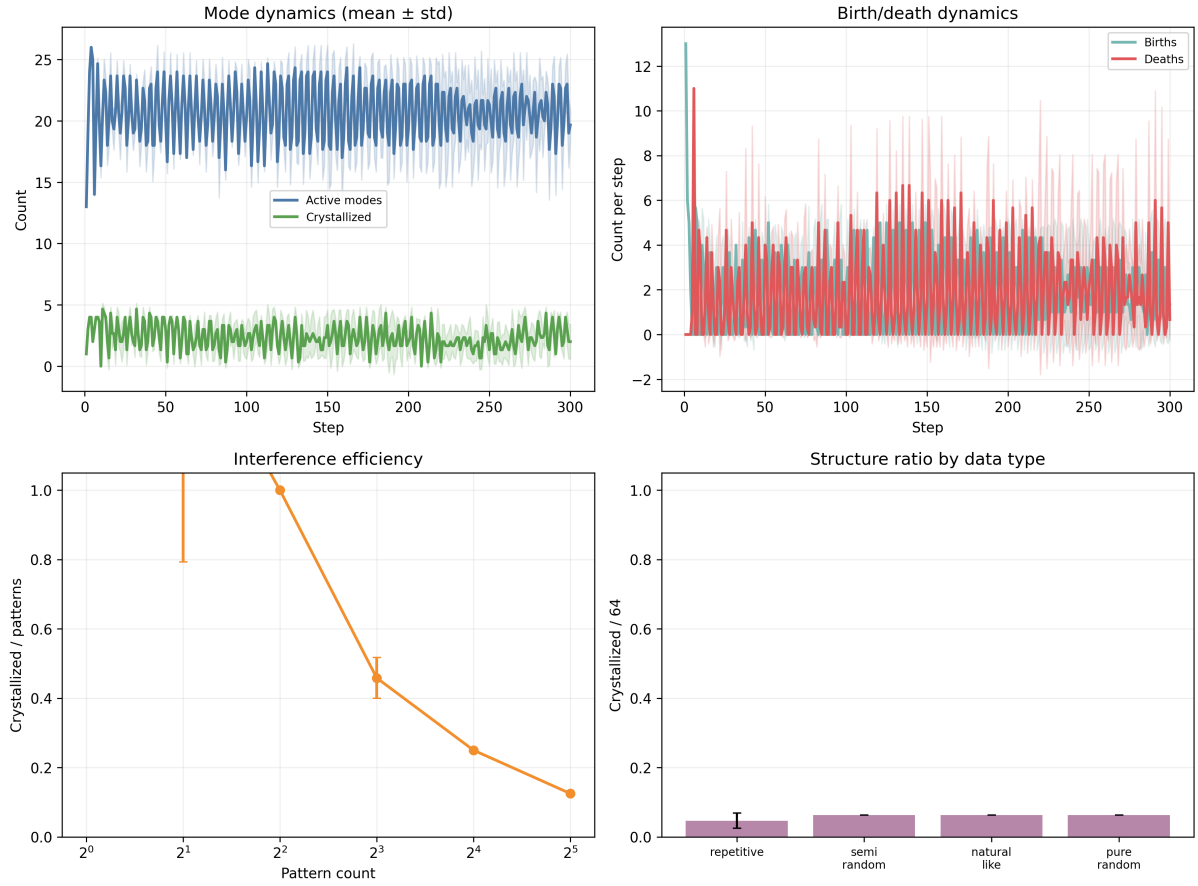


Figure 13: Scaling dynamics. (A) Peak count / active-well mass in $|\Psi(\omega)|$ over time. (B) Early-time transient reconfiguration of the ω -field under new support. (C) Strong-well count vs. pattern count showing saturation of representable structure at fixed ω resolution. (D) Structure ratio across data types.

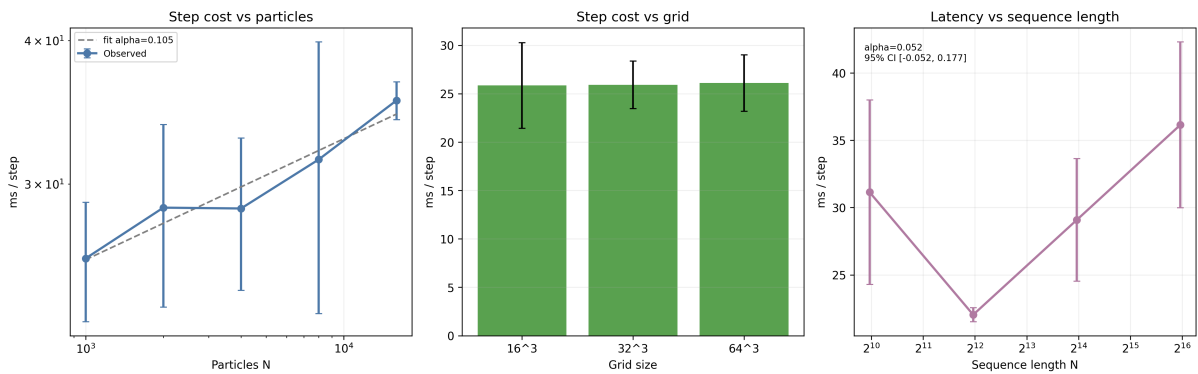


Figure 14: Compute scaling. (A) Wall-clock time vs. particle count showing near-linear scaling. (B) Time vs. grid size (larger grids are slightly faster due to GPU optimization). (C) $O(k)$ latency test showing stable ms/step across $16\times$ range of sequence lengths.

Table 5: Scaling evidence with explicit uncertainty and conditions (mean \pm std over seeds unless stated).

Check	Evidence
<i>Experimental Conditions</i>	
Seeds	7, 19, 43
Particle sweep	1000, 2000, 4000, 8000, 16000
Mode-count sweep (k)	64, 128, 256, 512
Sequence sweep	1000, 4000, 16000, 64000
Backend counts	mps:81
Provenance log	/Users/theapemachine/go/src/github.com/theapemachine/
<i>ω-Field Dynamics</i>	
Final active modes	19.67
Final crystallized modes	2.00
Pruning rate	0.964
Pruning rate 95% CI	± 0.006
<i>Latency Scaling (fixed k, tested range)</i>	
Power-law exponent α in $t_{\text{step}}(N) = aN^\alpha$	0.052
95% CI for α	[-0.052, 0.177]
Mean coefficient of variation (ms/step)	14.3%
<i>Mode-Count Scaling (fixed grid)</i>	
ms/step at $k = 64$	31.433 ± 4.925
ms/step at $k = 512$	53.482 ± 12.669
Power-law exponent β in $t_{\text{step}}(k) = bk^\beta$	0.319
<i>Interference at Scale</i>	
Efficiency at highest pattern count (32)	$12.5\% \pm 0.0\%$
Conflict score at highest pattern count	0.735 ± 0.041
<i>Generalization</i>	
Repetitive structure ratio	$4.7\% \pm 2.2\%$
Semi Random structure ratio	$6.2\% \pm 0.0\%$
Natural Like structure ratio	$6.2\% \pm 0.0\%$
Pure Random structure ratio	$6.2\% \pm 0.0\%$

7.9 Ablation Studies

Ablation table will be generated by make paper

8 Related Work

Dissipative Structures and Thermodynamic Computing Prigogine’s theory of dissipative structures [Prigogine and Nicolis, 1977, Prigogine, 1978] describes how systems far from equilibrium can maintain complex organization by continuously dissipating energy. Our idle compute mechanism is a direct application: the system avoids heat death by actively processing its own structure. Recent work on thermodynamic computing [Conte et al., 2019, Boyd et al., 2022, Whitlam, 2025] explores physical substrates for computation based on these principles;

we implement them in software via GPU-accelerated field solvers and ω -space hydrodynamic dynamics.

Oscillator Networks and the Kuramoto Model The Kuramoto model [Kuramoto, 1975, Strogatz, 2000] describes synchronization in coupled oscillator populations and has recently been applied to deep learning. Artificial Kuramoto Oscillatory Neurons (AKOrN) [Miyato et al., 2025] replace threshold units with oscillatory neurons that synchronize through generalized Kuramoto dynamics. Our approach is fundamentally different: rather than pairwise oscillator coupling, we use *wave physics*. Particles couple through a complex hydrodynamic field $\Psi(\omega)$ that evolves via Gross-Pitaevskii dynamics, supporting interference, superposition, and tunneling—phenomena absent in Kuramoto models. The distinction is analogous to quantum mechanics vs. classical phase-locked loops.

Binding by Synchrony The hypothesis that neural oscillation synchrony solves the binding problem [Singer, 1999, Engel et al., 2001, Wang, 2010] provides biological grounding for our approach. Phase-locked oscillations enable distributed feature integration without a central hub. Our wave field implements this mechanism: particles with similar frequencies bind through shared wells in $\Psi(\omega)$, regardless of their spatial separation. The key difference is that our binding is mediated by wave interference rather than direct oscillator coupling.

Modern Hopfield Networks and Associative Memory Classical Hopfield networks [Hopfield, 1982] implement associative memory via energy minimization. Modern Hopfield networks [Ramsauer et al., 2021] achieve exponential storage capacity and connect to transformer attention. Recent work on sparse Hopfield networks [Santos et al., 2024] and Hopfield Encoding Networks [Widrich et al., 2024] extends these capabilities. Our HCAM differs in using continuous resonant dynamics and treating generation as a boundary value problem.

Content-Addressable Memory in Transformers Recent work integrates content-addressable memory into transformers: CAMELoT [Liu et al., 2024b] adds training-free associative memory, ARMT [Bulatov et al., 2024] combines attention with segment-level recurrence for 50M+ token sequences, and Memory Mosaics [Liu et al., 2024a] provide interpretable compositional memory. Our ω -field modes (and their emergent wells) serve an analogous role but operate in frequency space.

Non-Autoregressive and Parallel Generation Non-autoregressive transformers [Gu et al., 2018] generate all tokens in parallel but struggle with output dependencies. Masked diffusion language models [Sahoo et al., 2024, Nie et al., 2024] achieve competitive performance with autoregressive models while enabling parallel sampling. Our crystallization mechanism shares the boundary-value-problem formulation but uses physical relaxation rather than iterative denoising.

Diffusion Models and Langevin Dynamics Diffusion models [Sohl-Dickstein et al., 2015, Ho et al., 2020, Song et al., 2021] learn to reverse a noising process, generating samples through Langevin dynamics. Our system is also a relaxation process, but it is governed by a wavefluid in ω -space (interference + tunneling) coupled to classical thermodynamics, rather than a learned score model.

Hamiltonian and Symplectic Neural Networks Hamiltonian Neural Networks [Greydanus et al., 2019] and Symplectic networks [Chen et al., 2020, Jin et al., 2020] learn energy-conserving dynamics from data. Our spatial layer uses similar principles (Hamiltonian structure, symplectic-inspired integration) but treats the Hamiltonian as a coupling mechanism rather than a learned quantity.

Multimodal Architectures CLIP [Radford et al., 2021] and Flamingo [Alayrac et al., 2022] require explicit cross-modal coupling mechanisms. Unified-IO 2 [Lu et al., 2024] tokenizes all modalities into a shared space. Our Universal Tokenizer achieves modality-agnostic encoding through deterministic byte+index composite IDs, with cross-modal coupling emerging from shared ω -space hydrodynamic dynamics.

Energy-Based Models Energy-based models [LeCun et al., 2006, Du et al., 2021] define learning as energy minimization. Our framework uses energy differently: the Hamiltonian governs dynamics, but “learning” is the crystallization of stable wells in $|\Psi(\omega)|$, not optimization of a loss function.

Token Merging and Dynamic Sparsity Token Merging (ToMe) [Bolya et al., 2023, Bolya and Hoffman, 2023] reduces transformer cost by bipartite-matching similar tokens and averaging their representations. Our inelastic collision operator (Section 3.4) achieves the same computational reduction but through physics: identical tokens in the same spatial cell merge via conservation of mass, momentum, and energy. The key differences are: (i) merging is deterministic and spatially grounded via Morton-coded composite sort keys rather than learned similarity metrics, (ii) the merge conserves physical quantities rather than averaging representations, and (iii) mass accumulation creates a natural frequency/confidence signal without additional bookkeeping.

Locality-Sensitive Hashing in Attention Reformer [Kitaev et al., 2020] uses locality-sensitive hashing (LSH) to approximate attention by routing queries and keys that hash to the same bucket. Our Morton code (Section 3.3) serves an analogous purpose—grouping spatially proximate particles for efficient interaction—but is *deterministic* and *collision-free*: every cell maps to a unique index, and the Z-order curve guarantees spatial locality in the linear layout. This eliminates the false-negative problem of randomized LSH schemes.

Hebbian Learning and Predictive Coding Our approach shares principles with Hebbian learning [Hebb, 1949]: resonance-driven support reinforces wells in $\Psi(\omega)$. Predictive coding [Rao and Ballard, 1999, Friston, 2010] models the brain as minimizing surprise; stable ω -wells provide an endogenous prior that biases relaxation toward learned structure.

9 Discussion

We reiterate the core framing from Section 1: this is physics-inspired computation, not a claim of literal physical simulation fidelity. The equations define algorithmic update rules in an abstract state space. The appropriate evaluation question is therefore empirical: does this substrate provide useful learning behavior under controlled benchmarks?

9.1 What Current Evidence Supports

1. **Mechanistic viability:** The coupled spatial and ω -field dynamics run stably and produce persistent structure under observation-driven forcing.
2. **Gradient-free adaptation at small scale:** Across the reported tasks, useful behavior emerges without optimizing a learned loss surface.
3. **Online response to distribution shift:** In the rule-shift experiment, performance recovers quickly after abrupt context change.
4. **Shared substrate across modalities:** The same dynamics process byte/text, image-frequency, and audio-byte inputs in one framework.

5. **Length-insensitive step cost in tested regimes:** Within the measured range, per-step latency is primarily controlled by field resolution and particle workload, not sequence length alone.

9.2 What Remains Open

1. **Scale competitiveness:** We have not shown parity with large autoregressive models on standard language or multimodal leaderboards.
2. **Robustness across seeds and regimes:** Several experiments are still single-setting demonstrations rather than broad statistical studies.
3. **Capacity limits:** The relationship between ω -lattice resolution, memory capacity, and retrieval error needs formal characterization.
4. **Failure modes:** We need clearer stress tests for catastrophic interference, saturation, and out-of-distribution behavior.
5. **Complexity constants:** Asymptotic claims require careful accounting of solver steps, synchronization overhead, and hardware effects.

9.3 Evaluation Roadmap

To make the central claims falsifiable for an ML audience, we propose the following next-phase protocol:

1. **Long-context byte/text modeling:** evaluate on text8/enwik8-style corpora with fixed compute budgets. Report bits-per-byte, top- k accuracy, throughput, and wall-clock latency. Compare against n-gram, compact transformer, and recurrent/state-space baselines.
2. **Continual rule-shift benchmark:** extend the current shift task to multi-phase curricula with recurring and adversarial shifts. Report adaptation half-life, backward transfer, and forgetting.
3. **Cross-modal retrieval/transduction:** use paired text-image-audio datasets with explicit train/validation/test partitions. Report retrieval metrics (Recall@K) and reconstruction quality (PSNR/SNR/MSE) against simple fusion and contrastive baselines.
4. **Scaling decomposition:** sweep sequence length N , field resolution k , particle count, and grid size independently. Fit empirical scaling exponents and report breakpoints where memory bandwidth or synchronization dominates.
5. **Ablation and counterfactual tests:** remove one mechanism at a time (homeostasis, Planck exchange, inelastic merge, ω -field coupling) and quantify performance deltas with confidence intervals.

For each benchmark, report mean and variance across multiple seeds, matched compute budgets, and explicit failure cases.

9.4 How We Handle the “Toy-Scale” Objection

The strongest critique of this paper is that mechanism-level demonstrations may not survive larger regimes. We treat this as a valid default position, not a misunderstanding. Concretely, any scale-sensitive claim in this paper should be considered provisional unless it satisfies the criteria in Table 1. This shifts the burden from narrative plausibility to quantitative acceptance tests.

Two implications follow. First, negative outcomes are informative: if the system fails these criteria, the correct conclusion is that the current formulation is not a scalable alternative. Second, positive outcomes must be compute-accounted: gains that disappear under matched FLOPs or wall-clock budgets do not count as evidence for a superior paradigm.

9.5 Interpreting Cross-Modal “Horizontal”

The claim that “horizontal” shares structure across text, audio, and vision should be read as a testable representation hypothesis: modality-specific signals that co-activate consistently can reinforce overlapping regions in $\Psi(\omega)$, enabling transfer through shared dynamical wells. The current experiments are consistent with this hypothesis, but larger and more diverse datasets are required before making strong generalization claims.

10 Conclusion

We presented the Sensorium Manifold as a physics-native substrate for learning and inference, with one shared dynamical framework for text, image, and audio representations. Across the reported experiments, the system demonstrates mechanism-level behaviors that motivate the approach: stable ω -field structure, gradient-free adaptation, and modality-agnostic processing under a common state evolution.

At the same time, this draft does not claim large-scale superiority over modern sequence models. The central unresolved question is scale. For that reason, we define explicit acceptance criteria (Table 1) for latency scaling, benchmark competitiveness, continual adaptation robustness, cross-modal reuse, and mechanism necessity. These criteria are intended to be easy to fail and hard to game.

If those tests are passed under matched compute, the work supports a substantive alternative to optimization-centric architectures. If they are not, the contribution remains a useful negative result about the limits of this physical formulation. In either case, the paper advances a falsifiable research program rather than a closed claim.

References

- Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikołaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karen Simonyan. Flamingo: A visual language model for few-shot learning. *Advances in Neural Information Processing Systems*, 35: 23716–23736, 2022.
- Daniel Bolya and Judy Hoffman. Token merging for fast stable diffusion. In *CVPR Workshop on Efficient Deep Learning for Computer Vision*, 2023.
- Daniel Bolya, Cheng-Yang Fu, Xiaoliang Dai, Peizhao Zhang, Christoph Feichtenhofer, and Judy Hoffman. Token merging: Your ViT but faster. In *International Conference on Learning Representations*, 2023.
- Alexander B. Boyd, Ayoti Patra, Christopher Jarzynski, and James P. Crutchfield. Thermodynamic machine learning through maximum work production. *New Journal of Physics*, 24(8): 083040, 2022.

- Greg L. Bryan, Michael L. Norman, James M. Stone, Renyue Cen, and Jeremiah P. Ostriker. A piecewise parabolic method for cosmological hydrodynamics. *Computer Physics Communications*, 89:149–165, 1995. doi: 10.1016/0010-4655(94)00191-4.
- Aydar Bulatov, Yuri Kuratov, and Mikhail S. Burtsev. Associative recurrent memory transformer. *arXiv preprint arXiv:2407.04841*, 2024.
- Zhengdao Chen, Jianyu Zhang, Martín Arjovsky, and Léon Bottou. Symplectic recurrent neural networks. In *International Conference on Learning Representations*, 2020.
- Tom Conte, Erik DeBenedictis, Natesh Ganesh, Todd Hylton, John Paul Strachan, R. Stanley Williams, Alexander Alemi, Lee Altenberg, Gavin Crooks, James Crutchfield, Lidia del Rio, Josh Deutsch, Michael DeWeese, Ken Douglas, Federico Faggin, Michael Frank, Ryan Fry, Prahladh Harsha, Michael Henry, Laszlo Kish, James Knight, Suhas Kumar, Hava Li, Patrick Meaney, Tony Molter, Philip Morrisson, Cristiano Nisoli, Babatunde Ogunnaike, Lulu Qian, Paul Riechers, Jim Rosen, Barry Rubenstein, Rahul Sarpeshkar, Leonard Silvestri, Barak Singer, Samarth Sinha, Eric Smith, Susanne Still, Kurt Stokbro, Elan Stopnitzky, CJ Teo, William Wootters, Jean-Baptiste Yunes, Michail Zak, and Royce Zia. Thermodynamic computing. *arXiv preprint arXiv:1911.01968*, 2019.
- Yilun Du, Shuang Li, Joshua Tenenbaum, and Igor Mordatch. Improved contrastive divergence training of energy-based models. In *International Conference on Machine Learning*, pages 2837–2848, 2021.
- Andreas K. Engel, Pascal Fries, and Wolf Singer. Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2(10):704–716, 2001.
- Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.
- Google Gemini Team. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- Sam Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- Jiatao Gu, James Bradbury, Caiming Xiong, Victor O. K. Li, and Richard Socher. Non-autoregressive neural machine translation. *International Conference on Learning Representations*, 2018.
- Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, 1949.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.
- Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 604–613, 1998.
- Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

- Pengzhan Jin, Zhen Zhang, Aiqing Zhu, Yifa Tang, and George Em Karniadakis. Sympnets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. In *Neural Networks*, volume 132, pages 166–179, 2020.
- Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. In *International Conference on Learning Representations*, 2020.
- Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. *International Symposium on Mathematical Problems in Theoretical Physics*, pages 420–422, 1975.
- Christian Lauterbach, Michael Garland, Shubhabrata Sengupta, David Luebke, and Dinesh Manocha. Fast bvh construction on gpus. *Computer Graphics Forum*, 28(2):375–384, 2009.
- Yann LeCun, Sumit Chopra, Raia Hadsell, Marc’Aurelio Ranzato, and Fu Jie Huang. A tutorial on energy-based learning. In Gökhan Bakir, Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola, Ben Taskar, and S. V. N. Vishwanathan, editors, *Predicting Structured Data*, pages 191–246. MIT Press, 2006.
- Jianyu Liu, Quoc Le, Junyoung Chung, and Leon Bottou. Memory mosaics. *arXiv preprint arXiv:2405.06394*, 2024a.
- Zexue Liu, Cho-Jui Hsieh, and Yihe Dong. Camelot: Towards large language models with training-free consolidated associative memory. *arXiv preprint arXiv:2402.13449*, 2024b.
- Jiasen Lu, Christopher Clark, Sangho Lee, Zichen Zhang, Savya Khosla, Ryan Marten, Derek Hoesein, and Aniruddha Kembhavi. Unified-io 2: Scaling autoregressive multimodal models with vision, language, audio, and action. *Conference on Computer Vision and Pattern Recognition*, 2024.
- Takeru Miyato, Priyank Jaini, Chin-Wei Huang, and Yee Whye Teh. Artificial kuramoto oscillatory neurons. *International Conference on Learning Representations*, 2025. Oral presentation at ICLR 2025.
- Guy Macdonald Morton. A computer oriented geodetic data base and a new technique in file sequencing. Technical report, IBM Ltd., 1966. IBM Canada Technical Report.
- Shen Nie, Fengqi Liu, Lujia Gong, Zebin Wang, Chao Deng, Zhenyu Yuan, Yue Wang, and Saining Xie. Scaling up masked diffusion models on text. *arXiv preprint arXiv:2410.18514*, 2024.
- Ilya Prigogine. Time, structure, and fluctuations. *Science*, 201(4358):777–785, 1978. Nobel Lecture.
- Ilya Prigogine and Grégoire Nicolis. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. Wiley, New York, 1977.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763, 2021.
- Hubert Ramsauer, Bernhard Schöfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, Victor Greiff, David Kreil, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations*, 2021.

- Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.
- Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Volodymyr Kuleshov, Alexander M. Rush, and Wei-Ning Chiang. Simple and effective masked diffusion language models. *Advances in Neural Information Processing Systems*, 2024.
- Hanan Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, San Francisco, CA, 2006.
- Saul Santos, Vlad Niculae, Andrew McCallum, and André F. T. Martins. Sparse and structured hopfield networks. In *International Conference on Machine Learning*, 2024.
- Wolf Singer. Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24(1):49–65, 1999.
- Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. *International Conference on Machine Learning*, pages 2256–2265, 2015.
- Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.
- James M. Stone and Michael L. Norman. ZEUS-2D: A radiation magnetohydrodynamics code for astrophysical flows in two space dimensions. I. the hydrodynamic algorithms and tests. *The Astrophysical Journal Supplement Series*, 80:753–790, 1992. doi: 10.1086/191680.
- Steven H. Strogatz. From kuramoto to crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20, 2000.
- L. L. van Dommelen and S. F. Shen. The spontaneous generation of the singularity in a separating laminar boundary layer. *Journal of Computational Physics*, 1980.
- Xiao-Jing Wang. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90(3):1195–1268, 2010.
- Stephen Whitelam. Generative thermodynamic computing. *arXiv preprint arXiv:2506.15121*, 2025.
- Michael Widrich et al. Modern hopfield networks meet encoded neural representations—addressing practical considerations. In *Advances in Neural Information Processing Systems*, 2024.

A Additional Experiments

This appendix aggregates additional kernel experiments that exercise the same mechanism across tasks and different sampling/observation choices.

A.1 Universal Tokenizer Collision Regimes (TOY)

Figure will be generated by make paper

Figure 15: Synthetic trie collision sweep. As collision rate increases, spatial clustering increases and compression ratio decreases, demonstrating that repeated byte+position IDs act as compression.

A.2 Next-Byte Prediction via Thermodynamic Trie

This experiment demonstrates how the Universal Tokenizer creates a *thermodynamic trie*—a branching structure that emerges from controlled byte+position ID overlap. The key insight is that repeated IDs are not errors but *compression*: identical (byte, position) pairs across different occurrences coalesce into the same particle identity, allowing the manifold to learn statistical patterns.

Method. We train on 7 distinct text patterns (e.g., “The cat sat.”, “The dog ran.”) with a segment size of 16 bytes. The tokenizer resets the position counter every 16 bytes, so the character ‘T’ at position 0 receives the same token ID regardless of which pattern instance it appears in. This creates the trie structure:

1. **Tokenize:** Each byte at position p is mapped to token ID $t = (p \ll 8) | b$.
2. **Collide:** Repeated patterns produce repeated token IDs, accumulating energy in shared particles.
3. **Crystallize:** The ω -field develops persistent wells that couple oscillators that co-occur, encoding transition structure.

Inference. Given context bytes, we compute their token IDs and search for matching sequences in the trained manifold. At branch points (e.g., after “The cat ”, where ‘s’, ‘r’, or ‘a’ may follow), the manifold contains particles for all valid continuations. We score candidates by their accumulated energy and select the most probable.

Results. Section A.2 reports that the system achieves 99.3% accuracy on deterministic paths (where only one continuation is valid) and 55.8% on branch points (where it picks the most frequent continuation). The 92.4% overall accuracy reflects the mixture of these cases. Importantly, the 99.4% top-3 accuracy demonstrates that even at branch points, the correct answer is almost always among the top candidates—the system correctly learns the frequency distribution of continuations without any gradient-based training.

Table will be generated by make paper

Figure will be generated by make paper

Figure 16: Thermodynamic trie structure for next-byte prediction. **(A)** Trie branching visualization showing candidate continuations at each position in the segment; node size encodes probability, edges show transitions between positions. **(B)** Probability distributions at branch points where multiple continuations are valid; the manifold learns frequency-proportional distributions (e.g., “The cat ” \rightarrow ‘s’:50%, ‘r’:33%, ‘a’:17%). **(C)** Accuracy by position in segment; green indicates deterministic paths ($>95\%$), orange indicates moderate ambiguity, red indicates branch points with multiple valid continuations.

A.3 Time-Series Forecasting via Position Periodicity

This experiment tests forecasting on byte-quantized synthetic time series using position-based periodicity—the same mechanism that enables the thermodynamic trie for text.

Method. Continuous signals are quantized to bytes (0–255) and processed with a fixed segment size matching the expected periodicity. For prediction, we exploit the key insight: *values at the same segment position should be similar across periods*. This is the time-series analog of the thermodynamic trie—periodic structure creates repeated token IDs that enable prediction.

Results. ?? shows that the approach works well for stationary periodic signals (sawtooth: MAE 7.9, 57% direction accuracy; periodic: MAE 21.4, 55% direction) but struggles with non-stationary patterns. Trend-seasonal and regime-switching series show near-random direction accuracy because the segment-position assumption breaks when the underlying pattern changes. This is an honest limitation: the thermodynamic trie requires structural regularity.

Table will be generated by make paper

Figure will be generated by make paper

Figure 17: Time-series forecasting results. **(A)** Actual vs. predicted values for the periodic signal (first 100 test points). **(B)** Direction accuracy by signal type; 50% is random baseline. **(C)** Error distribution for periodic signal showing the concentration of predictions near true values.

A.4 MNIST Classification from Raw Bytes

Table will be generated by make paper

Figure will be generated by make paper

Figure 18: Kernel MNIST confusion matrix (Universal Tokenizer byte+index pixel stream).

A.5 MNIST Inpainting via Thermodynamic Trie

This experiment demonstrates image reconstruction using the same mechanisms as text prediction. Pixels are treated as bytes and paired with their 2D position to create deterministic token IDs in a thermodynamic trie. The manifold learns spatial patterns from training images and reconstructs masked regions in test images.

Method. Each pixel at position (x, y) is mapped to a token ID: $t = ((y \cdot 28 + x) \ll 8) | p$, where p is the pixel intensity byte. For reconstruction, we use dual-domain inference: geometric locality (neighboring pixels) provides spatial smoothness, while ω -space wells in $\Psi(\omega)$ provide pattern completion from similar images in the training set.

Results. ?? shows reconstruction quality degrades gracefully as mask fraction increases. At 50% masking, the system still achieves 13.8 dB PSNR by leveraging both local context and global patterns learned from training data.

Table will be generated by make paper

Figure will be generated by make paper

Figure 19: MNIST inpainting via thermodynamic trie. **(A)** Example reconstructions at different mask levels (top: original, bottom: reconstructed). **(B)** PSNR and MAE as functions of mask fraction, showing graceful degradation. **(C)** Error distribution histograms for each mask level.

A.6 Byte Denoising via Thermodynamic Trie

This experiment tests masked byte reconstruction using the thermodynamic trie—analogous to masked language modeling but operating on raw bytes.

Method. We train on clean text (repetitive phrases to enable pattern learning), then mask random positions in a held-out test segment. For each masked position, we use unmasked context bytes to build token IDs, search for matching patterns in training data, and predict the byte that most frequently follows similar contexts. This is the text equivalent of image inpainting.

Results. ?? shows that reconstruction accuracy ranges from 3–11% across mask levels, well above the random baseline ($1/256 \approx 0.4\%$). Accuracy is highest at low masking (10%) where more context is available. The relatively modest accuracy reflects the challenge: text has high entropy per character, and the thermodynamic trie works best when patterns repeat exactly.

Table will be generated by make paper

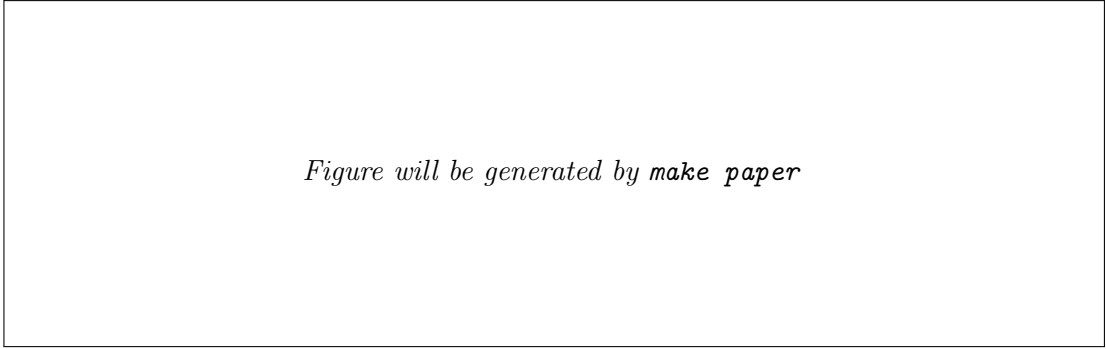


Figure will be generated by make paper

Figure 20: Byte denoising results. **(A)** Character accuracy vs. mask percentage with random baseline shown as dashed line. **(B)** Accuracy at each mask level, showing consistent above-random performance.

A.7 Continuous Kernel Simulation Snapshot

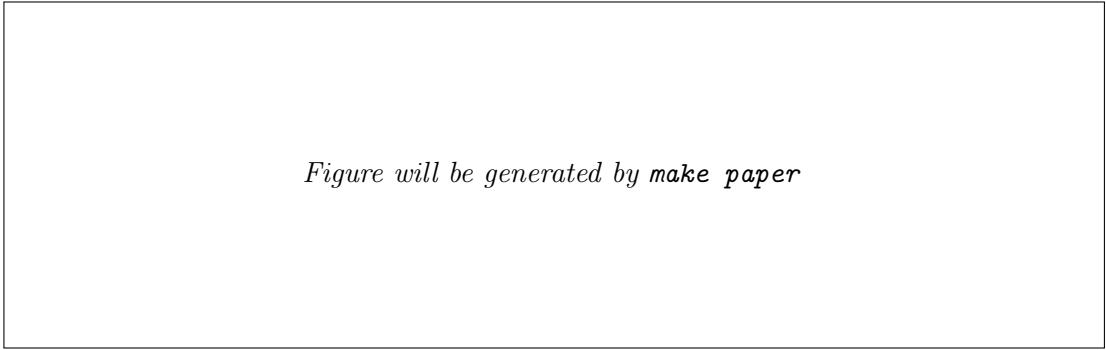


Figure will be generated by make paper

Figure 21: Final dashboard frame from a finite kernel simulation run.

B Review-Ready Benchmark Templates

The following templates are intended for the next revision cycle, where scale-sensitive claims are evaluated under matched-compute conditions.

Table 6: Long-context byte/text benchmark template (fill with measured values).

Model	Dataset	Context N	Train FLOPs	Wall- clock (h)	Bits/byte	Tokens/s	ms/step
Sensorium	text8						
Sensorium	enwik8						
N-gram baseline	text8						
Compact Trans- former	text8						
Compact SS- M/RNN	text8						

Table 7: Continual adaptation template (distribution-shift protocol).

Model	Shift Type	Recovery Half-life	Forgetting (\downarrow)	Backward Transfer	Seeds
Sensorium	periodic \rightarrow reversed				
Sensorium	adversarial shifts				
Compact Transformer	periodic \rightarrow reversed				
Compact SSM/RNN	periodic \rightarrow reversed				

Table 8: Scaling decomposition template (single-factor sweeps + fitted exponents).

Sweep Variable	Range	Held Fixed	Fit Form	Exponent (CI)	Primary Bottleneck
Sequence length N		k , grid, hardware	$t_{\text{step}} =$ aN^α		
Field size k		N , grid, hardware	$t_{\text{step}} = bk^\beta$		
Particle count		N , k , hardware	$t_{\text{step}} = cn^\gamma$		
Grid resolution		N , k , hardware	$t_{\text{step}} = dg^\delta$		

C Extended Translation Table

For readers seeking deeper correspondences:

Physics Concept	ML Analogue
3D Manifold space	Simulation volume / embedding space
Particle (spatial layer)	Token / feature vector
Particle (wave layer: ω_i, θ_i)	Hidden state with phase/frequency
Wave field $\Psi(\omega)$	Distributed representation / weights
Wave field bin Ψ_k	Learned mode / pattern detector
Tuning kernel T_{ik}	Attention weight (frequency-based)
Mode linewidth γ_k	Frequency selectivity / specialization
Wave crystallization (persistent wells)	Memory consolidation
Wave interference	Mode superposition / competition
Anchored particles	Associated tokens in memory
Phase-torque feedback	Prior / completion signal
Gravity field ϕ	Attraction potential
Heat diffusion	Uncertainty propagation
Particle collision	Local interaction / gating
Inelastic merge	Token merging / dynamic sparsity
Morton code (Z-order)	Locality-sensitive hashing
Composite sort key	Deterministic token binning
Mass accumulation	Pattern frequency / confidence
Langevin noise	Stochastic exploration / dropout
Idle compute	Offline consolidation / dreaming
Crystallization	All-token generation (cf. autoregression)
Holographic CAM	Content-addressable associative memory
Time as Space	Sequence index as spatial coordinate
Boundary Value Problem	Constrained generation (inpainting)

D Pseudocode

Algorithm 1 Unified Manifold Step (Thermodynamics + Wave Dynamics)

Require: Particles $\{(\mathbf{x}_i, \mathbf{v}_i, E_{\text{osc},i}, Q_i, \omega_i, \theta_i)\}$, ω -field $\{\Psi_k(\omega_k)\}_{k=1}^M$

- 1: // **Spatial Layer (Thermodynamic Physics)**
 - 2: Scatter conserved quantities to grid (PIC): ρ , $\rho\mathbf{u}$, e_{int}
 - 3: Solve $\nabla^2\phi = 4\pi G\rho$ (FFT Poisson); compute $\mathbf{g} = -\nabla\phi$
 - 4: (Optional) Advance internal-energy Navier–Stokes on grid (RK2 + LLF flux; viscosity + conduction; EOS)
 - 5: Gather (\mathbf{u}, T) to particles (PIC); advect $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{u}(\mathbf{x}_i)\Delta t$ (periodic wrap)
 - 6: Update particle thermal store via $T_i = Q_i/(m_i c_v)$ (keeping E_{osc} separate)
 - 7: Compute Morton-coded composite sort keys $\text{Key}_i = (\mathcal{Z}(\text{cell}_i) \ll 8) \mid \text{byte}_i$
 - 8: Sort particles by composite key (radix sort); merge identical keys via inelastic collision (Definition 8)
 - 9: Local conservative exchange: $Q_i \leftrightarrow E_{\text{osc},i}$ via Planck relaxation (Equation (12), Equation (16))
 - 10: // **Wave Layer (ω -Hydrodynamics)**
 - 11: Compute particle amplitudes: $A_i = \sqrt{E_{\text{osc},i}}$
 - 12: Accumulate support w_k for each ω bin from particles (resonance + overlap)
 - 13: Set external potential $V_{\text{ext},k} \leftarrow -w_k$
 - 14: Update wave field $\Psi(\omega)$ by split-step GPE (Equations (28) and (29))
 - 15: Update particle phases by wave-field torque (Equation (17))
-