The Sensorium Manifold:

Native Multimodality via Isomorphism

Daniel Owen van Dommelen
Independent Research - WORKING DRAFT
theapemachine@gmail.com

February 3, 2026

Abstract

We introduce the *Sensorium Manifold*, a thermodynamic computing substrate that replaces the autoregressive paradigm with global energy minimization. Current AI models rely on serial token generation and backpropagation-based optimization. We propose a system governed by *Hamiltonian dynamics*, where data is represented as a field of coupled oscillators and learning is the emergence of resonant modes (carriers).

We introduce the *Universal Tokenizer*, a modality-agnostic input mechanism that maps raw bytes and sequence indices to unique attractor IDs via hashing. This treats all data—text, image, audio—as a single branching stream where "collision is compression." Structure emerges not through architectural bias, but through the physical bifurcation of energy flows.

The system operates via three principles: (1) **Spectral Entanglement**, where distant oscillators couple via shared carrier frequencies; (2) **Metabolic Gating**, where carriers persist only if energetically maintained by resonance; and (3) **Crystallization**, a parallel inference process where inputs are "dumped" into the manifold and allowed to relax into a minimum-energy configuration, filling in missing information simultaneously rather than causally. We demonstrate that this system conserves energy (symplectic integration), adapts to rule shifts online, and enables massive parallelism impossible in Transformer architectures.

A Note on Vocabulary. This paper presents a computational substrate based on Hamiltonian mechanics and coupled oscillators. 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
Oscillator	Input Token	Has phase/frequency; exists in continuous time.
Carrier (Soliton)	Hidden State / Weight	A standing wave that couples oscillators.
Hamiltonian (H)	Loss Function	Conserved quantity; system minimizes potential V.
Spectral Coupling	Attention Mechanism	Non-local entanglement via frequency resonance.
Crystallization	Inference	Global parallel relaxation, not serial generation.
Metabolism	Regularization	Carriers decay if they do not receive energy.
Universal Tokenizer	Embedding Layer	Deterministic hashing of raw bytes; no training.
Phase Locking	Pattern Matching	Information encoded in relative phase angles.
Symplectic Integrator	Optimizer	Preserves energy phase-space; no gradient descent.

1 Introduction

The dominant paradigm in machine learning treats computation as optimization: define a loss function, compute gradients via backpropagation, and descend toward minima. This has proven remarkably effective, yet it imposes constraints that may not reflect how physical systems learn. Biological neural networks do not have access to global error signals; they adapt through local interactions governed by thermodynamic and biochemical principles.

We propose an alternative paradigm: *thermodynamic computation*. In physical systems, structure emerges from energy flow, entropy production, and homeostatic regulation. We apply these principles to construct a learning system where:

- **Particles** represent activated concepts, carrying energy through continuous space
- **Attractors** are stable points in the space, corresponding to learned representations
- **Bonds** encode relationships between concepts, with mass that grows from use and decays from disuse
- **Heat** captures uncertainty and accumulated noise, driving exploration
- **Homeostasis** regulates system activity through adaptive baselines, preventing runaway excitation or quiescence

1.1 Native Multimodality

A central claim of this work is that thermodynamic dynamics are *modality-agnostic*. Current multimodal architectures—CLIP [?], Flamingo [?], Gemini [?]—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, bonding, metabolism, 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 $\pi_1 : \mathcal{F}_1 \to \mathbb{R}^D$ and $\pi_2 : \mathcal{F}_2 \to \mathbb{R}^D$ 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, carriers couple them automatically. The manifold dynamics remain unchanged.

1.2 Contributions

1. A complete thermodynamic framework for learning without backpropagation (Section 2)
2. The Sensorium Manifold: a unified substrate for native multimodality (Section 6)
3. Sparse bond graph dynamics that avoid dense $V \times V$ matrices (Section 7)
4. Dissipative structure maintenance through stochastic traversal (Section 8)
5. Empirical validation on cross-modal transduction and adaptation (Section 10)

2 Thermodynamic Framework

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.

2.1 The Particle-Attractor System

The Sensorium Manifold $\mathcal{S}$ is a flat Euclidean space $\mathbb{R}^D$. Within this space, we distinguish two types of entities:

Definition 1 (Particle)**.** *A particle p_i is a transient entity with:*

- *Position* $\boldsymbol{x}_i \in \mathbb{R}^D$
- *Energy* $E_i \geq 0$
- *Heat* $Q_i \geq 0$

- *Modality tag* $\mu_i \in \{text, audio, image, \ldots\}$

Particles are created by encoders, evolve through dynamics, and are consumed by decoders.

Definition 2 (Attractor). *An attractor* a_j *is a persistent entity with:*

- *Position* $\boldsymbol{a}_j \in \mathbb{R}^D$
- *Energy* $E_j \geq 0$
- *Heat* $Q_j \geq 0$
- *Optional: associated token, frequency, or other modality-specific data*

Attractors represent learned structure. Particles diffuse toward nearby attractors.

2.2 Thermodynamic Quantities

Definition 3 (Total Energy). *The total energy of the system is:*

$$E_{total} = \sum_i E_i + \sum_j E_j + \sum_i \varepsilon_i \tag{1}$$

where ε_i *is the excitation (active energy available for flow).*

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)*
- *Flow through bonds (friction)*
- *Conflict between competing predictions*

Heat drives diffusion (exploration) and decays homeostatically.

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:*

$$\rho = \frac{\log(1 + E_{total})}{\log(1 + \mathcal{B}) + \varepsilon} \tag{2}$$

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

$$\mathcal{B}_{t+1} = (1 - \alpha)\mathcal{B}_t + \alpha\, E_{total}, \quad \alpha = \frac{\Delta t}{\tau + \Delta t} \tag{3}$$

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 [**?**]. 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 Diffusion Dynamics

Particles diffuse toward nearby attractors. The dynamics are:

$$\frac{d\boldsymbol{x}_i}{dt} = -\nabla_{\boldsymbol{x}_i} U(\boldsymbol{x}_i) + \sqrt{2DQ_i}\,\boldsymbol{\xi}(t) \tag{4}$$

where $U(\boldsymbol{x})$ is a potential defined by attractor positions:

$$U(\boldsymbol{x}) = -\sum_j E_j \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{a}_j\|^2}{2\sigma^2}\right) \tag{5}$$

and $\boldsymbol{\xi}(t)$ is white noise. Heat drives the stochastic term, enabling exploration when uncertainty is high.

Remark 2 (Discretization)**.** *In practice, we use Euler-Maruyama discretization with timestep Δt. We do not claim to solve continuous-time SDEs exactly.*

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 hash function:

$$\mathrm{ID} = \mathcal{H}(\mathrm{Byte}, \mathrm{Index}) \pmod N \tag{6}$$

where Byte $\in [0, 255]$ is the raw data and Index is the sequence position (or spatial coordinate).

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

- **Convergence**: All inputs sharing a prefix flow into the same initial attractors.
- **Bifurcation**: When the data diverges (e.g., Image A has a white pixel at $(0, 1)$, Image B has black), the energy flow splits.

This creates a *Thermodynamic Trie*. The system naturally learns the topology of the data stream by observing which transitions (bonds) are energetically favorable.

3.2 Modality Agnosticism

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

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

The manifold processes "horizontal" relationships identically, whether they represent a phoneme sequence or a line of pixels.

4 Hamiltonian Dynamics

Unlike dissipative neural networks, the Resonant 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 Resonant Ghost 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 standing wave potential that permeates the space. We term this the *Resonant Ghost Field*.

4.1.1 Field Definition

The Ghost Field Ψ is the aggregate scalar potential generated by the population of M crystallized carriers (solitons). Unlike a neural network weight which exists to multiply a specific input, the Ghost Field exists as a background potential even in the absence of particles.

The Hamiltonian interaction term is defined as:

$$V_{\text{ghost}}(\mathbf{q}) = -\sum_{k=1}^{M} \underbrace{\mu_k}_{\text{Mass}} \cdot \left| \sum_{i=1}^{N} \exp\left(-\frac{(\omega_i - \Omega_k)^2}{2\sigma_k^2} + i(\theta_i - \phi_k) \right) \right|^2 \tag{7}$$

where:

- μ_k is the metabolic mass (importance) of Carrier k.
- Ω_k is the intrinsic frequency of the Carrier.
- ω_i, θ_i are the frequency and phase of the input oscillator i.

4.1.2 Action at a Distance (Wormholes)

This potential creates a non-Euclidean geometry. In the raw sequence index, token i (at $t = 0$) and token j (at $t = 1000$) are distant. However, if both tokens resonate with Carrier k (i.e., $|\omega - \Omega_k| < \sigma_k$), the potential V_{ghost} creates a deep energy well that binds them.

Effectively, the Resonant Ghost Field folds the manifold, creating a *semantic wormhole*. Two oscillators tuned to the same carrier are entangled with zero distance in phase space, regardless of their separation in sequence space.

4.1.3 The "Ghost" Property

We term this field a "Ghost" field because it represents *potential energy*. A crystallized carrier for the concept "Cat" exists in the manifold with $\mu_k > 0$. It is invisible and consumes no active compute (kinetic energy) until an input oscillator with the hash for "Cat" enters the system.

Upon entry, the oscillator immediately falls into the potential well of the Ghost Field, transferring kinetic energy into the carrier. The carrier "wakes up" (resonates), and through the entanglement described above, immediately pulls the "Meow" oscillator (if present) or hallucinates it (if missing) via the shared potential well.

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 $\frac{d\mathcal{H}}{dt} \approx 0$.

$$p(t + \Delta t/2) = p(t) - \nabla V(q(t))\frac{\Delta t}{2} \tag{8}$$

$$q(t + \Delta t) = q(t) + \frac{p(t + \Delta t/2)}{m}\Delta t \tag{9}$$

$$p(t + \Delta t) = p(t + \Delta t/2) - \nabla V(q(t + \Delta t))\frac{\Delta t}{2} \tag{10}$$

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

4.3 Metabolic Gating

While the short-term dynamics are conservative, the long-term structure is dissipative. Carriers (memory units) are subject to metabolic decay:

$$\frac{d\text{Mass}_k}{dt} = \text{Income}_k - \text{Cost}(\bar{E}) \tag{11}$$

Carriers that successfully resonate with input data receive "Income" (energy injection). Carriers that fail to resonate starve and vanish. This implements continuous, online model selection (Occam's Razor) via thermodynamics.

5 Crystallization

Autoregressive models generate data serially ($t_1 \to t_2 \to t_3$). The Resonant Manifold performs inference via *Global Relaxation* or *Crystallization.*

5.1 The Bucket Dump

We slice the input data (e.g., an image with missing patches) into buckets and inject them into the manifold simultaneously.

1. **Excitation**: Injected oscillators vibrate at their intrinsic frequencies.
2. **Resonance**: Stored carriers (learned patterns) that match the partial input begin to resonate sympathetically.
3. **Hallucination**: The resonating carriers pump energy into the missing oscillators (the gaps).

5.2 Phase Locking

The solution emerges when the system settles into a phase-locked state (a local energy minimum). This is a non-causal process: future tokens can stabilize past tokens, and boundary conditions propagate inward. This allows for massive parallelism, as the entire sequence "crystallizes" out of the noise at once.

6 The Sensorium Manifold

6.1 Unified Particle Space

The key insight is that particles from different modalities can coexist in a shared embedding space if we project their native coordinates appropriately.

Definition 6 (Spectral Encoding)**.** *For a sensory input with native spectral representation $\boldsymbol{f} \in \mathcal{F}$, the spectral encoding is:*

$$\boldsymbol{x} = \pi(\boldsymbol{f}) \in \mathbb{R}^D \tag{12}$$

where $\pi : \mathcal{F} \to \mathbb{R}^D$ is a projection that preserves distance relationships.

For modalities with native dimension $d < D$, we pad with zeros:

$$\pi(\boldsymbol{f}) = [\boldsymbol{f}; \mathbf{0}_{D-d}] \tag{13}$$

For modalities with native dimension $d > D$, we use random projection (Johnson-Lindenstrauss):

$$\pi(\boldsymbol{f}) = \boldsymbol{f} \cdot \boldsymbol{P}, \quad \boldsymbol{P} \in \mathbb{R}^{d \times D}, \quad \boldsymbol{P}_{ij} \sim \mathcal{N}(0, 1/D) \tag{14}$$

6.2 Modality-Specific Encoders

6.2.1 Audio Encoding

Audio is encoded via 1D Fourier transform:

$$\hat{x}(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} \, dt \tag{15}$$

Each frequency bin f_k with magnitude $|\hat{x}(f_k)|$ becomes a particle with:

- Position: $\pi([f_k]) \in \mathbb{R}^D$
- Energy: $E_k = |\hat{x}(f_k)| / \sum_j |\hat{x}(f_j)|$
- Phase: $\phi_k = \arg(\hat{x}(f_k))$ (stored for reconstruction)

6.2.2 Image Encoding

Images are encoded via 2D Fourier transform:

$$\hat{I}(u, v) = \int \int I(x, y) e^{-2\pi i (ux + vy)} \, dx \, dy \tag{16}$$

Each spatial frequency (u_k, v_k) becomes a particle with:

- Position: $\pi([u_k, v_k]) \in \mathbb{R}^D$
- Energy: $E_k = |\hat{I}(u_k, v_k)| / \sum_j |\hat{I}(u_j, v_j)|$
- Phase: $\phi_k = \arg(\hat{I}(u_k, v_k))$

6.2.3 Text Encoding

Text tokens are encoded via learned or pretrained embeddings:

$$\boldsymbol{x}_k = \text{embed}(t_k) \in \mathbb{R}^D \tag{17}$$

If using a semantic embedding of dimension d, we project to $\mathbb{R}^D$ via π.

6.3 The Unified Dynamics

Once all particles are in $\mathbb{R}^D$, the dynamics are identical:

1. **Diffusion**: Particles drift toward attractors and diffuse based on heat
2. **Bonding**: Co-activated particles form bonds (see Section 7)
3. **Metabolism**: Bonds grow from use, decay from cost
4. **Homeostasis**: System energy is regulated against adaptive baseline

The manifold does not "know" that a particle came from audio vs. image. It only sees positions, energies, and heats. Cross-modal relationships emerge naturally when particles from different modalities co-activate.

6.4 Modality-Specific Decoders

To output in a specific modality, we select particles by modality tag and reconstruct:

6.4.1 Audio Decoding

Collect all audio-tagged particles. Reconstruct the spectrum:

$$\hat{x}(f_k) = E_k \cdot e^{i\phi_k} \tag{18}$$

Apply inverse FFT to obtain the waveform.

6.4.2 Image Decoding

Collect all image-tagged particles. Reconstruct the 2D spectrum:

$$\hat{I}(u_k, v_k) = E_k \cdot e^{i\phi_k} \tag{19}$$

Apply inverse FFT2D to obtain the image.

7 Sparse Bond Graph Dynamics

Relationships between particles are encoded in a sparse directed graph, not dense weight matrices.

7.1 Bond Graph Representation

Definition 7 (Bond Graph)**.** *The bond graph $\mathcal{G} = (V, E)$ is a sparse directed graph where:*

- *V is the set of attractor indices*
- *$E \subseteq V \times V$ is the set of directed bonds*
- *Each bond $(i, j) \in E$ has mass $m_{ij} \geq 0$ and trace $\tau_{ij} \geq 0$*

The graph is stored as a sparse edge list (I, J, M, T), scaling with the number of observed transitions, not $O(V^2)$.

7.2 Bond Formation (Hebbian Dynamics)

Bonds form and strengthen through co-activation:

Principle 2 (Hebbian Bonding)**.** *When particles at attractors i and j are co-activated (e.g., observed in sequence), the bond (i, j) receives mass:*

$$m_{ij} \leftarrow m_{ij} + \Delta t \cdot \varepsilon_i \cdot \varepsilon_j \tag{20}$$

For sequential data, we observe $i \to j$ and inject mass:

$$m_{ij} \leftarrow m_{ij} + \Delta t \cdot \Pi \cdot (1 + \text{depletion}_i) \tag{21}$$

where Π is the metabolic pressure (normalized surprise) and depletion_i is accumulated from sink encounters during stochastic traversal.

7.3 Metabolism

Active bonds undergo metabolism: they receive income from use and pay cost for existence.

Definition 8 (Bond Income)**.**

$$income_{ij} = \varepsilon_i \cdot \tilde{m}_{ij} \cdot \frac{1}{1 + \bar{Q} + \varepsilon} \tag{22}$$

where $\tilde{m}_{ij} = m_{ij} / \sum_k m_{ik}$ *is the normalized mass and* $\bar{Q}$ *is mean heat.*

Definition 9 (Bond Cost)**.**

$$cost_{ij} = \rho \cdot \bar{\varepsilon} \cdot \frac{m_{ij}}{\bar{m} + \varepsilon} \tag{23}$$

where ρ *is the homeostatic ratio.*

The mass update is:

$$m_{ij} \leftarrow \max(0, m_{ij} + \Delta t \cdot (\mathrm{income}_{ij} - \mathrm{cost}_{ij})) \tag{24}$$

This is not gradient descent. There is no loss function. Bonds that are used grow; bonds that are unused decay. The homeostatic ratio modulates the rate.

7.4 Energy Flux

Energy flows through bonds according to the flux equation:

$$\Phi_j = \sum_{i:(i,j) \in E} \varepsilon_i \cdot \tilde{m}_{ij} \tag{25}$$

This is the amount of excitation arriving at attractor j from all sources. The computation is sparse—we only sum over existing bonds.

7.5 Adaptive Pruning

Rather than fixed thresholds, bonds are pruned when their mass falls below the mean outgoing mass from their source:

$$\mathrm{prune}(i,j) \iff m_{ij} < \frac{1}{|N_{\mathrm{out}}(i)|} \sum_{k \in N_{\mathrm{out}}(i)} m_{ik} \tag{26}$$

This maintains sparsity while adapting to the scale of each source.

8 Stochastic Traversal

Between observations, the system performs stochastic traversal—sampling paths through its bond graph under thermal noise. This is not passive relaxation toward equilibrium, but active dissipation: the system expends energy to explore its own structure.

Following Prigogine's theory of dissipative structures [?], we recognize this as the mechanism by which the system maintains organization far from equilibrium. The traversal *exports entropy* (by exploring and abandoning unproductive paths) while *reducing internal entropy* (by discovering shortcuts that sharpen future predictions). Without this process, the system would approach heat death—maximum entropy, no gradients, no useful work possible.

This is not "offline training"—there is no training data. The system processes its own structure.

8.1 Transitive Closure

If bonds $A \to B$ and $B \to C$ exist, we can infer a shortcut $A \to C$:

$$m_{AC} = \Delta t \cdot \tilde{m}_{AB} \cdot \tilde{m}_{BC} \cdot \exp\left(-\frac{\|\boldsymbol{a}_A - \boldsymbol{a}_C\|^2}{\bar{d}^2}\right) \tag{27}$$

The geometric factor $\exp(-d^2/\bar{d}^2)$ discounts shortcuts between distant attractors.

8.2 Conflict Resolution

Sources with high outgoing entropy (ambiguous predictions) receive noise injection:

$$\varepsilon_i \leftarrow \varepsilon_i + \Delta t \cdot \mathcal{N}(0, \sigma_i^2), \quad \sigma_i \propto S_i \tag{28}$$

where $S_i = -\sum_j \tilde{m}_{ij} \log \tilde{m}_{ij}$ is the entropy of outgoing bonds. Subsequent metabolism allows competition to sharpen predictions.

8.3 Stochastic Trajectories

The system samples hypothetical trajectories from its current structure:

Algorithm 1 Stochastic Trajectory Sampling

budget $\leftarrow E_{\text{dream}} \cdot (1 + \gamma \cdot \max(0, \rho - 1))$
while budget > 0 **do**
    Sample starting attractor i proportional to ε_i
    path_prob $\leftarrow 1$
    for $t = 1, \ldots, T_{\max}$ **do**
        Compute $P(\cdot|i)$ from bond graph
        if P is empty (sink) **then**
            depletion$_i \leftarrow$ depletion$_i + \Delta t$ ▷ Mark sink
            break
        end if
        Sample $j \sim P$
        path_prob $\leftarrow$ path_prob $\cdot P(j|i)$
        Reinforce (i, j) with mass $\propto$ path_prob
        $i \leftarrow j$
        budget $\leftarrow$ budget $- (1 + Q_i/\bar{Q})$
    end for
end while
depletion $\leftarrow$ depletion $\cdot \exp(-\Delta t \cdot \rho / \overline{\text{depletion}})$ ▷ Homeostatic decay

Depletion accumulates at sinks (states with no outgoing flux). When real observations arrive, depletion modulates plasticity—the system is more receptive to learning transitions from depleted attractors. This is not a biological hunger signal; it is a local measurement of incomplete structure that increases the rate of bond formation.

9 Cross-Modal Transduction

The Sensorium Manifold enables cross-modal transduction without architectural changes. We describe the mechanism.

9.1 Carrier Coupling

To transduce between semantic (text) and spectral (audio/image) representations, we introduce a population of *carriers*—particles that have positions in both semantic and spectral spaces.

Definition 10 (Carrier)**.** *A carrier c_k has:*

- *Semantic position $\boldsymbol{s}_k \in \mathbb{R}^D$*
- *Spectral position $\boldsymbol{f}_k \in \mathbb{R}^{d_{spec}}$ (1D for audio, 2D for image)*
- *Energy $E_k \geq 0$*
- *Heat $Q_k \geq 0$*

Carriers learn to couple semantic and spectral spaces through co-activation. When semantic particle at $\boldsymbol{s}$ and spectral particle at $\boldsymbol{f}$ are observed together, nearby carriers receive energy:

$$E_k \leftarrow E_k + \Delta t \cdot w_{\text{sem}}(\boldsymbol{s}, \boldsymbol{s}_k) \cdot w_{\text{spec}}(\boldsymbol{f}, \boldsymbol{f}_k) \tag{29}$$

where w are kernel functions (e.g., Gaussian).

9.2 Forward Transduction (Semantic to Spectral)

To transduce semantic input to spectral output:

1. Compute carrier activations from semantic input
2. Each carrier projects its energy to its spectral position
3. Collect spectral energy distribution
4. Decode via inverse FFT

9.3 Backward Transduction (Spectral to Semantic)

The same mechanism works in reverse. The manifold is bidirectional.

10 Experiments

We validate the framework on three tasks:

10.1 Rule-Shift Adaptation

We evaluate adaptation to distributional shifts using a controlled benchmark:

- **Forward phase** (steps 1-1000): The sequence "The cat sat on the mat" repeats
- **Reverse phase** (steps 1001-2000): The sequence reverses

The system must rapidly unlearn forward transitions and learn reverse transitions, online, without gradient-based retraining.

10.2 Cross-Modal Audio Synthesis

We demonstrate semantic-to-audio transduction using the carrier coupling mechanism. Given text tokens, the system produces spectral distributions that can be synthesized to audio.

Table 1: Rule-shift experiment results. Run `make paper` to generate.

Metric	**Value**
Pre-shift accuracy	–
Post-shift accuracy (immediate)	–
Post-shift accuracy (recovered)	–
Steps to 80% recovery	–
Final chunk count	–

Figure will be generated by `make paper`

Figure 1: Rule-shift adaptation dynamics. The system recovers to baseline accuracy within [X] steps after complete reversal of sequential structure.

10.3 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.

10.4 Ablation Studies

11 Related Work

Dissipative Structures Prigogine's theory of dissipative structures [??] describes how systems far from equilibrium can maintain complex organization by continuously dissipating energy. Our stochastic traversal mechanism is a direct application: the system avoids heat death by actively exporting entropy through path exploration, while reducing internal entropy by discovering structural shortcuts.

Thermodynamic Computing Recent work explores physical substrates for computation based on thermodynamic principles [Conte et al., 2019]. We draw inspiration from these ideas but implement them in software.

Hebbian Learning Our approach shares principles with Hebbian learning [Hebb, 1949]. We extend this with thermodynamic regulation and sparse graph dynamics.

Energy-Based Models Energy-based models [LeCun et al., 2006] define learning as energy minimization. Our framework differs: energy is for homeostatic regulation, not optimization.

Multimodal Architectures CLIP [?] uses contrastive learning to align image and text encoders. Flamingo [?] introduces cross-attention layers to fuse visual features into language

Table 2: Ablation study. Run `make paper` to generate.

Condition	Pre-shift Acc.	Post-shift Acc.	Recovery Steps
Full system	–	–	–
No hierarchy	–	–	–
No pondering	–	–	–
No homeostasis	–	–	–

models. These approaches require explicit coupling mechanisms that must be designed per modality pair. We propose that cross-modal coupling can emerge from shared dynamics: particles coexist in a common space, and Hebbian co-activation creates associations without architectural intervention.

Predictive Coding Predictive coding [Rao and Ballard, 1999] models the brain as a prediction engine minimizing surprise. Our surprise-driven plasticity relates to this framework.

12 Discussion

12.1 What We Claim

1. **Native multimodality**: All sensory modalities can be processed by the same thermodynamic dynamics on a shared manifold.
2. **No backpropagation**: Learning emerges from local, Hebbian-style dynamics regulated by homeostasis.
3. **Online adaptation**: The system adapts continuously to streaming data and distributional shifts.
4. **Dissipative self-organization**: Stochastic traversal maintains structure far from equilibrium by exporting entropy while discovering shortcuts.

12.2 What We Do Not Claim

1. **Strict thermodynamics**: We use thermodynamic *metaphors*. Energy is not conserved; there is no detailed balance.
2. **Gradient-free optimization**: We are not optimizing a loss function without gradients. We sidestep optimization entirely.
3. **Transformer replacement**: Our experiments are on small-scale tasks. We make no claims about scaling to language model pretraining.
4. **Continuous dynamics**: Our "continuous" dynamics are Euler-discretized. We do not solve PDEs exactly.

12.3 The Physics of "Horizontal"

We claimed that "the physics of horizontal is the same whether it is a word, a sound wave, or a pixel pattern." Let us make this precise:

- **Text**: The word "horizontal" is a token with a D-dimensional embedding.

- **Audio**: A sound panning left-to-right has specific frequency characteristics (Doppler, stereo phase).
- **Image**: A horizontal line has energy concentrated at $v \approx 0$ in the 2D frequency domain.

All three representations enter the manifold as particles. Through co-activation during training, carriers couple these representations. The word “horizontal” activates carriers that also respond to horizontal image frequencies and horizontal audio characteristics. This is not a metaphor—it is the mechanism.

13 Conclusion

We have presented the Sensorium Manifold, a unified thermodynamic substrate for native multimodal computation. By representing all sensory inputs as spectral distributions in a shared Euclidean space, we achieve modality-agnostic dynamics. Learning emerges from local thermodynamic interactions—energy flow, Hebbian bonding, homeostatic regulation—without backpropagation.

The framework suggests an alternative to the optimization-centric paradigm of modern machine learning. Physical principles—thermodynamics, diffusion, homeostasis—may offer paths to adaptive systems that are better suited to continuous, online, embodied learning.

The physics of “horizontal” really is the same across modalities. And that, perhaps, is how perception should work.

References

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 Morrision, 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.

Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory.* Wiley, New York, 1949.

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.

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.

A Extended Translation Table

For readers seeking deeper correspondences:

Physics Concept	ML Analogue
Euclidean space $\mathbb{R}^D$	Embedding space
Particle diffusion	Gradient flow / Langevin dynamics
Attractor basin	Class region / prototype
Energy injection	Forward pass activation
Heat dissipation	Regularization / dropout (sort of)
Homeostatic ratio ρ	Adaptive learning rate
Bond mass m	Edge weight in sparse graph
Eligibility trace τ	Temporal credit (no backprop)
Metabolic cost	Weight decay / L2 regularization
Metabolic income	Hebbian update
Transitive closure	Shortcut / skip connections
Stochastic trajectories	Monte Carlo sampling / rollouts
Depletion signal	Curiosity / exploration bonus
Sink	Absorbing state / dead end
Dissipative structure	Self-organizing system
Spectral distribution	Frequency representation
Carrier coupling	Cross-modal alignment
Event horizon locality	Local / sparse attention

B Hyperparameter Settings

Table 3: Hyperparameter settings for all experiments.

Parameter	Symbol	Value
Integration timestep	Δt	0.02
Homeostasis time constant	τ	1.0
Numerical epsilon	ε	10^{-8}
Dream energy budget	–	32.0
Dream sampling temperature	–	1.0
Carrier homeostasis τ	τ_c	5.0
Thermal resistance	R_{th}	1.0
Viscosity	μ	1.0

C Pseudocode

Algorithm 2 Unified Manifold Step

Require: Particles $\{(\boldsymbol{x}_i, E_i, Q_i)\}$, Attractors $\{(\boldsymbol{a}_j, E_j)\}$, Bond graph $\mathcal{G}$

```
Project all particles to common space: z_i = π(x_i)
Compute ρ via Equation (2)
Diffusion: Update particle positions via Equation (4)
Bonding: For co-activated pairs, inject mass via Equation (20)
Metabolism: Update bond masses via Equation (24)
Flux: Propagate energy via Equation (25)
Heat: Q_i ← Q_i + Δt · |Φ_i|/Φ̄
Damping: ε_i ← ε_i · exp(−Δt · ρ/ε̄)
Pruning: Remove bonds below per-source mean
Sync energy and heat back to particles
```