Codes as Coordination: How Coupled Protocells Generate Symbolic Communication Before Genes

Ian Todd

Sydney Medical School

University of Sydney

Sydney, NSW, Australia

`itod2305@uni.sydney.edu.au`

Abstract

We propose that symbolic codes first emerged as coordination interfaces between coupled protocellular compartments, not as transmitted messages requiring pre-existing encoder-decoder machinery. Using a minimal model of weakly coupled chemical compartments with high-dimensional internal dynamics and multi-channel boundary signaling, we show that discrete, reproducible symbol sequences arise spontaneously from coordination equilibria. The key mechanism is substrate competition (lateral inhibition), which discretizes continuous dynamics without external logic. Our simulations show that networks of 61 coupled compartments reliably map 32 environmental configurations to 32 unique 4-symbol sequences with near-perfect reproducibility (mean 98.4% across 20 random seeds), and a receiver colony with identical dynamics reaches well-separated attractor states enabling 99% decoding accuracy. We argue that this "codes as coordination" framework addresses the apparent paradox of abiogenesis: information does not arise *de novo* from chemistry, but rather emerges when coherent coordination dynamics collapse into reproducible invariants. Communication precedes information storage. The genetic code is not the origin of biological meaning—it is the compression of meaning that first existed in the distributed coupling fields between cells.

Keywords: origin of life; genetic code; protocells; coordination; bioelectric signaling; reservoir computing; game theory

1 Introduction: The Code Emergence Problem

The origin of the genetic code remains one of the deepest puzzles in biology. How did discrete, symbolic information arise from continuous chemical processes? Standard accounts assume that codes require encoders, messages, and decoders—but this raises a circularity problem: how can encoding machinery exist before the code that specifies it?

We propose a resolution: **codes first emerged as coordination interfaces between coupled protocellular compartments**. They were not transmitted messages but stable patterns that allowed neighbors to influence each other's behavior without full state knowledge.

This reframing dissolves the circularity. The "code" is not something one compartment sends to another. It is the shared equilibrium that emerges when coupled systems settle into mutually compatible states. The encoder-message-decoder structure we recognize in mature biological systems is a *late refinement*—the internalization and compression of coordination protocols that first existed in the distributed coupling fields between cells.

This work builds on recent results establishing fundamental limits on observation and control in high-dimensional biological systems. Todd [29] showed that measurement thresholds create irreducible uncertainty in sub-Landauer regimes; Todd [30] demonstrated that timing constraints limit accessible information in continuous substrates; and Todd [31] proved that when high-dimensional dynamics (H_{dyn}) exceed observer channel capacity (C_{obs}), stable codes *must* form as a thermodynamic necessity. The present paper provides a concrete physical demonstration of this code formation mechanism.

1.1 Outline

Section 2 develops the theoretical framework: codes as coordination equilibria, the group-first hypothesis, and the evolutionary sequence from coupled dynamics to genetic compression. Section 3 addresses prebiotic plausibility, arguing that lipid vesicle networks are the minimal substrate satisfying all requirements for code emergence. Section 4 presents our physical model and simulation results. Section 5 discusses implications for abiogenesis, artificial life, and the philosophy of biology.

2 Theoretical Framework

2.1 Codes Are Coordination Equilibria

Consider a network of weakly coupled protocells:

- Each protocell has high-dimensional internal dynamics (many coupled molecular species)

- Neighbors communicate through **multiple channels**: bioelectric (ion gradients, membrane potential), chemical (pH, metabolites), mechanical (membrane tension), and redox signaling

- These signals are naturally discretized by threshold physics (voltage-gated transitions, precipitation, phase changes)

- Selection favors configurations where neighbors reliably respond to each other's boundary states

This multi-channel architecture provides redundancy, cross-validation, and temporal separation (fast electrical, slow chemical) [34, 35]. Levin and colleagues have demonstrated that

bioelectric patterns carry morphogenetic "codes" in biological development—voltage gradients that specify tissue identity [1, 2, 3]. We propose this is not a derived feature of complex life but a *primitive* capability: protocell networks coordinating via boundary signals before genetic encoding existed.

The resulting "code" is not a transmitted message. It is a **shared interface**—a stable pattern that allows coordination without requiring each compartment to know the other's full internal state.

Decoding as Free Energy Minimization. The "decoder" is not a separate apparatus— it is any neighboring compartment whose local phase space becomes constrained by the distributed coupling field. A compartment "reads" the code by minimizing its free energy relative to the boundary conditions imposed by its neighbors. The "meaning" of a symbol is the stability it induces in the receiver. Successful decoding is the transition to a coordination equilibrium.

2.2   The Group-First Hypothesis

A lone protocell faces a brutal tradeoff:

- Too open → mixes with environment, loses identity

- Too closed → cannot exchange resources, stagnates

- Too complex → collapses into monopoly attractors

A **network of weakly coupled compartments** relaxes these constraints:

1. **Spatial heterogeneity becomes memory**: Different compartments occupy different internal states, but the network stabilizes a shared macrostate

2. **Weak coupling enables coordination without collapse**: Too little coupling produces chaos; too much produces synchronization; intermediate coupling produces stable diversity [25, 26]

3. **Selection can operate at the group level**: Network-level "phenotypes" (stable symbol regimes) persist and outcompete others

This suggests early life was more likely a **coupled compartment network** than isolated protocells [7, 8, 9].

The Interface IS the Organism. Before internalization, "life" exists in the distributed coupling fields *between* compartments—in the geometry of coordination, not in any single cell. Selection pressure forces individual compartments to internalize compressed models of this distributed dynamic, enabling prediction of neighbor behavior. **The genetic code is the crystallization of the "between" into the "within."**

2.3   The Evolutionary Sequence

Codes do not appear all at once. We propose a developmental sequence:

Stage	What Exists	What "Code" Means
0	Coupled dynamics	No codes yet—just coordination
1	Stable boundary symbols	Coordination equilibria: "when I'm in state A, you respond with X"
2	Symbols label internal regimes	Proto-semantics: "I'm UV-tolerant," "I'm resource-hoarding"
3	Slow variables persist symbols	Message storage without genes—hysteresis, membrane composition
4	Compression into templates	Genetic code as late optimization

Key claim: You learn to send messages before you learn to store them. Communication is primary; information storage is secondary [22, 23, 24]. Genes are not the origin of codes; they are a compression of codes that already existed as coordination interfaces.

We hypothesize that there is no purely non-social form of life—that social game theory is not an emergent property of living systems but its foundation [4, 5, 6]. The deeper claim is stronger still: **nonliving systems can exhibit social dynamics**. Coupled oscillators "negotiate" stable phase relationships; competing catalysts "cooperate" via shared substrates; vesicle networks "coordinate" boundary states. These are not metaphors—they are the same game-theoretic structures that organize primate societies [37], now visible in chemistry because the math is substrate-independent. Life did not invent sociality; life is what happens when chemical sociality becomes heritable.

2.4   Coherence Is Not Information

A critical distinction underlies this framework:

Coherence	Information
High-dimensional phase order	Discrete symbols
Constraint, correlation, coordination capacity	Addressable, copyable, decodable tokens
The ability to move as a coupled whole	Something that can be reproduced across contexts
Not information	Information

The common category error in origin-of-life research is treating "structured dynamics" as "encoded symbols." They are not the same.

Coherence enables coordination. **Information** is what emerges when coherent dynamics collapse into reproducible invariants—stable tokens that can be transmitted, stored, and decoded.

The bridge: *codes form when coherence collapses into reproducible invariants.* This explains the apparent paradox: how can information arise from chemistry? Answer: it doesn't arise *de novo*. Chemistry first produces coherent coordination; codes emerge when that coordination stabilizes into discrete, copyable forms.

2.5 Intelligence as Ergodicity Defense

Why would coupled compartments evolve toward higher-dimensional dynamics? Because **high-dimensional systems are harder to collapse**.

Define **ergodicity defense** as the ability to maintain history-dependent, non-equilibrium macrostates under perturbation. A system that forgets its past (ergodic) cannot coordinate. A system that remembers (non-ergodic) can maintain identity while adapting.

Proposition: Intelligence, in its most primitive form, is ergodicity defense—the maintenance of non-ergodic dynamics against entropic collapse [27, 28, 31].

High effective dimensionality serves this goal:

- More internal directions to absorb perturbations without losing identity

- More reconfiguration paths for adaptation

- More orthogonal modes for coordination with neighbors

Selection pressure is not "make a code" or "maximize yield." It is: **avoid collapse into ergodicity while remaining responsive to environment and neighbors.**

3 Prebiotic Plausibility: Why Vesicles?

Our model posits networks of coupled protocellular compartments as the substrate for code emergence. This section addresses whether such structures could have existed prebiotically, and why vesicles are the most plausible candidate.

3.1 Evidence for Prebiotic Vesicle Formation

Lipid vesicles form spontaneously under conditions consistent with early Earth:

1. **Meteoritic delivery**: Amphiphilic molecules capable of forming membranes have been identified in carbonaceous chondrites, including the Murchison meteorite [11, 12]. These include fatty acids, alcohols, and polycyclic aromatic hydrocarbons with membrane-forming properties.

2. **Hydrothermal synthesis**: Simulations of hydrothermal vent conditions produce amphiphilic molecules from simple precursors (CO_2, H_2) via Fischer-Tropsch-type reactions [13, 14]. Alkaline vents provide both the chemical gradients and the confined spaces conducive to protocell formation [10, 15].

3. **Spontaneous assembly**: Above critical micelle concentrations, fatty acids spontaneously form bilayer vesicles in aqueous solution. Unlike phospholipids, short-chain fatty acids (C8–C12) form membranes that are permeable enough to allow exchange of small molecules while maintaining compartmentalization [16, 17].

4. **Growth and division**: Szostak and colleagues have demonstrated that fatty acid vesicles can grow by incorporating additional amphiphiles from the environment and divide under mild shear forces—without enzymatic machinery [18, 19]. This provides a plausible pathway from chemistry to self-reproducing compartments.

3.2 Alternative Compartmentalization Substrates

We acknowledge several alternative models for prebiotic compartmentalization:

Substrate	Advantages	Limitations
Coacervates	Form spontaneously from polymers; concentrate molecules; Oparin's original model	No defined boundary; less structured; unclear heredity mechanism
Hydrothermal vent pores	Mineral-catalyzed chemistry; natural gradients; fixed compartments	Geometry fixed by rock; cannot divide or evolve independently
Mineral surfaces	Catalytic activity; template-capable (clays); concentrate reactants	Two-dimensional; no encapsulation; limited to surface chemistry
Ice channels	Concentrate solutes via eutectic exclusion; stabilize polymers	Narrow temperature range; episodic availability
Phase-separated droplets	Membraneless organelles exist in modern cells; form via liquid-liquid separation	No clear prebiotic formation pathway; less stable than vesicles

Each substrate can provide compartmentalization, but they differ in whether they support the *full set* of requirements for code emergence via coordination.

3.3 Requirements for Coupled Compartment Dynamics

Our framework requires compartments that satisfy five conditions:

1. **Spatial separation**: Distinct interior volumes that can maintain different chemical states.

2. **Semi-permeability**: Boundaries that allow selective exchange of signals (ions, small molecules, electrical potentials) while retaining macromolecules and maintaining gradients.

3. **Internal chemistry**: Capacity to support non-equilibrium reaction dynamics—the "high-dimensional reservoir" that generates complex internal states.

4. **Neighbor coupling**: Physical contact or proximity enabling boundary-to-boundary signaling between adjacent compartments.

5. **Scalability**: Ability to form extended networks (arrays, chains, clusters) where collective dynamics can emerge.

Vesicles uniquely satisfy all five requirements:

- *Spatial separation*: Bilayer membranes create distinct aqueous compartments.

- *Semi-permeability*: Fatty acid membranes are permeable to small molecules and ions; membrane proteins (or prebiotic analogs) can create selective channels [16].

- *Internal chemistry*: Encapsulated reaction networks can operate far from equilibrium, driven by transmembrane gradients.

- *Neighbor coupling*: Vesicles in contact can exchange signals via membrane fusion, hemifusion, gap junction-like connections, or diffusion through shared aqueous phases. Bioelectric signaling (voltage gradients across membranes) provides a fast coordination channel [1].

- *Scalability*: Vesicles spontaneously form aggregates, chains, and two-dimensional arrays when concentrated on surfaces or in confined geometries [20, 21].

Coacervates satisfy (1), (3), and partially (2), but lack the structured boundaries needed for (4) and (5). Mineral pores satisfy (1)–(4) but fail (5)—they cannot form mobile, evolvable networks. Vesicles are the minimal prebiotic structure that supports the full coordination-to-code pathway.

3.4 The Mechanism Is Substrate-Agnostic

While we focus on vesicles for concreteness and experimental tractability, we emphasize that the *mechanism*—codes emerging from coordination equilibria between coupled high-dimensional systems—is substrate-agnostic.

Any system satisfying the five requirements above could, in principle, generate symbolic communication via the pathway we describe:

> Coupled dynamics $\rightarrow$ coordination equilibria $\rightarrow$ stable boundary symbols $\rightarrow$ compressed codes

This generality is important: it means the transition from chemistry to codes is not contingent on a specific molecular implementation but emerges from *geometric and dynamical constraints* on coupled compartments.

4 Physical Architecture and Simulation

4.1 Operational Definitions

Before describing the model, we define key terms operationally:

- **Symbol**: A discrete pattern of channel saturations across the N_{readout} output channels of a compartment. Specifically, a symbol is the binary sign pattern $\sigma = (\text{sign}(r_1), \ldots, \text{sign}(r_{N_{\text{readout}}}))$ where r_i is the i-th readout channel value after equilibration. This yields an alphabet of size up to $2^{N_{\text{readout}}}$, though substrate competition typically concentrates outputs into a much smaller effective alphabet.

- **Character**: A sequence of 4 symbols, one per temporal forcing cycle. Characters encode environmental configurations as temporal trajectories, not static snapshots.

- **Code**: The mapping from environmental configurations to characters. A code is "functional" if distinct inputs map to distinct outputs (no collisions) and the mapping is reproducible across trials.

- **Discretization**: The transformation from continuous internal dynamics to discrete symbols. In our model, discretization arises from substrate competition (lateral inhibition), not from externally imposed thresholds. The threshold at zero for sign extraction is physically interpretable as the reference electrode potential in differential measurement.

Note on alphabet size: With $N_{\text{readout}} = 30$ channels, the theoretical alphabet is $2^{30} \approx 10^9$ symbols. However, substrate competition forces winner-take-most dynamics that constrain the system to a small subset of saturated patterns. Our result—mapping 32 inputs to 32 unique outputs—demonstrates *reliable emergence of reproducible discrete tokens*, not a hard capacity-limit feat.

4.2   Model Equations

We model internal vesicle chemistry as a generic high-dimensional dissipative dynamical system. This is a *phenomenological* reservoir, not a literal reaction network—we do not claim a specific stoichiometry. The reservoir captures the essential property that protocells contain many coupled molecular species with complex, history-dependent dynamics.

Internal state dynamics: Each compartment i has internal state $\mathbf{x}_i(t) \in \mathbb{R}^{N_{\text{internal}}}$ evolving as:

$$\mathbf{x}_i(t+1) = (1-\alpha)\mathbf{x}_i(t) + \alpha \tanh\left(\mathbf{W}\mathbf{x}_i(t) + \mathbf{W}_{\text{in}}\mathbf{s}_i(t) + \mathbf{W}_{\text{couple}}\bar{\mathbf{r}}_{\mathcal{N}(i)}(t) + \mathbf{W}_{\text{fb}}\mathbf{r}_i(t)\right) \quad (1)$$

where α is the leak rate, $\mathbf{W}$ is the recurrent weight matrix (scaled to spectral radius $\rho < 1$), $\mathbf{s}_i(t)$ is the local stimulus, $\bar{\mathbf{r}}_{\mathcal{N}(i)}$ is the mean readout of neighbors, and $\mathbf{r}_i$ is the compartment's own readout.

Raw activity: The raw output activity is $\mathbf{a}_i = \max(0, \mathbf{W}_{\text{out}}\mathbf{x}_i)$.

Substrate dynamics: The shared substrate pool S_i evolves via:

$$\frac{dS_i}{dt} = k_{\text{rep}}(S_{\max} - S_i) - k_{\text{cons}} \sum_j a_{ij} \cdot \frac{S_i^h}{K_m^h + S_i^h} \quad (2)$$

where k_{rep} is the replenishment rate, k_{cons} is the consumption rate, h is the Hill coefficient (cooperativity), and K_m is the half-saturation constant.

Competitive allocation: Output channel j receives substrate fraction:

$$\phi_{ij} = \frac{a_{ij}^h}{\sum_k a_{ik}^h + \epsilon} \quad (3)$$

This is the quasi-steady-state (QSSA) solution to competitive binding. The denominator arises from mass conservation, not computational normalization.

Readout: The final readout is the mean-centered allocation scaled by substrate availability:

$$r_{ij} = G \cdot (\phi_{ij} - \bar{\phi}_i) \cdot \frac{S_i}{K_m + S_i} \tag{4}$$

where G is the measurement gain (instrumentation parameter) and $\bar{\phi}_i$ is the mean allocation. Mean-centering models differential measurement against a reference electrode—standard electrochemistry practice.

4.3   Parameter Table

Parameter	Description	Value
N_{vesicles}	Number of compartments	61
N_{internal}	Internal dimensions per compartment	128
N_{readout}	Output channels per compartment	30
ρ	Spectral radius of $\mathbf{W}$	0.92
α	Leak rate	0.25
S_{max}	Maximum substrate pool	10.0
k_{cons}	Consumption rate	0.5
k_{rep}	Replenishment rate	0.3
K_m	Half-saturation constant	1.0
h	Hill coefficient	2.0
G	Readout gain	20.0
N_{cycles}	Temporal forcing cycles	4
N_{steps}	Equilibration steps per cycle	70

Convergence: We use 70 equilibration steps followed by 20 averaging steps. Sensitivity analysis confirms results are stable for 50–100 equilibration steps.

4.4   System Components

We implement a minimal version of the coupled-compartment architecture:

Component	Function	Implementation
Compartments	Semi-autonomous agents	61 vesicles in hexagonal array
Fast dynamics	Internal chemistry	128 dimensions per compartment
Substrate competition	Emergent discretization	Lateral inhibition (30 channels)
Multi-channel coupling	Neighbor communication	Boundary signal exchange
Boundary readout	Symbol formation	Collective coordinates
Spatial heterogeneity	Symmetry breaking	Center-edge, top-bottom gradients

4.5   The Discretization Mechanism: Substrate Competition

The critical innovation is **substrate competition**—a lateral inhibition mechanism where output channels compete for shared metabolic resources. This creates emergent discretization: continuous dynamics collapse into discrete attractor basins without external logic.

The mechanism uses standard biochemistry:

- **Hill kinetics**: Saturation function with cooperativity ($n = 2$)

- **Michaelis-Menten dynamics**: Substrate consumption and replenishment

- **Competitive allocation**: Each channel's output $\propto \text{activity}^n / \sum(\text{activity}^n)$

This allocation formula is the **quasi-steady-state approximation (QSSA)** solution to competitive binding kinetics [36]. While the equation $a_i^h / \sum_j a_j^h$ mathematically resembles the "Softmax" function used in machine learning, it here arises from **conservation of mass**: the "sum" in the denominator happens *physically* because the substrate pool depletes. When channel A grabs a molecule, channel B cannot have it. This competitive exclusion forces winner-take-most dynamics that discretize continuous states into stable binary outputs without programmed thresholds.

4.6 Environmental Heterogeneity

Each vesicle experiences spatially varying stimulus conditions (center-edge light gradients, top-bottom temperature gradients, left-right flow gradients), mimicking the spatial structure of real prebiotic environments like tidal pools or hydrothermal vent systems. *Crucially, these represent non-informational geometric constraints*—a rock shading part of the pool, proximity to a heat source, differential ion exposure. The complexity is not in the stimulus; it is in the system's ability to differentiate continuous gradients into discrete coordination states. This breaks symmetry and increases the number of distinguishable coordination equilibria.

Additionally, **temporal forcing** creates genuine sequence structure. Each of 4 temporal cycles experiences different environmental conditions (modeling diurnal UV variation, tidal ion fluxes, temperature rhythms). This ensures the 4-symbol character sequence is a *genuine temporal trajectory*, not merely the same symbol repeated.

4.7 Simulation Results

Results for a single network (seed $= 42$):

Metric	Result
Compartments	61 vesicles (hexagonal array)
Internal dimensions per vesicle	128
Readout channels per vesicle	30
Unique input→character mappings	32/32 (no collisions)
Reproducibility (same network, multiple trials)	100%
Separation ratio (between/within)	335,361×
Env–Attractor correlation	0.72
Decoder accuracy	100%

Note: These results are for a single fixed network. Cross-seed robustness is reported in the following subsection.

Key findings:

1. **Complete discrimination**: 32 input configurations map to 32 unique output sequences with no collisions. The network has sufficient capacity for full discrimination.

2. **Emergent discretization**: 89% of readout values are saturated ($|x| > 0.5$). The chemistry does the digitizing, not external logic.

3. **Geometry preservation**: Environment–attractor correlation of 0.72 indicates that similar inputs map to similar outputs—the code preserves structure.

4. **Physics decoder**: A second vesicle array ("receiver colony") with identical dynamics achieves 100% decoding accuracy without machine learning.

4.8 Robustness Across Random Seeds

To rule out "lucky initialization," we tested the system across 20 independent random seeds (different initial conditions, weight matrices, and coupling topologies):

Metric	Mean $\pm$ SD	Range
Unique codes	31.2 ± 1.1	28–32
Separation ratio	$298,000 \pm 42,000$	210,000–380,000
Reproducibility	$98.4\% \pm 2.1\%$	94–100%
Decoder accuracy	$99.1\% \pm 1.3\%$	96–100%

All 20 seeds produce functional codes with ≥ 28 unique mappings. The *specific* symbol assignments vary (each seed discovers a different code), but the *structural properties* (discretization, reproducibility, decodability) are robust. This demonstrates that code emergence is a generic property of the architecture, not a special case.

4.9 The Receiver Colony and Decoding Accuracy

The decoder is a **second vesicle array** ("receiver colony") that receives the encoder's output signal:

- Receiver vesicles have identical dynamics to encoder vesicles but *different random internal structure* (different $\mathbf{W}$, $\mathbf{W}_{\text{in}}$, etc.)

- They receive the encoder's 20-dimensional code signal (center + edge aggregates)

- They process this signal through their own reservoir dynamics

- Each environmental configuration drives the receiver into a distinct attractor state

Decoding procedure: An *external observer* measures decoding accuracy by comparing receiver attractor states to canonical (mean) receiver patterns via nearest-neighbor matching:

$$\hat{c} = \arg \min_c \sum_{t=1}^{4} \|\mathbf{r}_{\text{rec}}(t) - \bar{\mathbf{r}}_c(t)\|^2 \tag{5}$$

where $\bar{\mathbf{r}}_c$ is the canonical receiver response to configuration c.

This satisfies the Shannon structure: Environment → Encoder → Code signal → Receiver → Distinguishable attractor. The receiver reaches 32 well-separated attractor states; the external observer merely *verifies* this separation. If a downstream "readout chemistry" were added inside the receiver (e.g., winner-take-all channel selection), decoding would be fully autonomous.

4.10   Manifold Expansion Validation

To validate the theoretical prediction that coupling increases the effective dimensionality of the output manifold, we measured the participation ratio $D_{\text{eff}} = (\sum \lambda_i)^2 / \sum \lambda_i^2$ across the coupling sweep $\kappa \in \{0, 0.05, 0.10, 0.15, 0.20, 0.30, 0.50\}$:

κ	D_{eff}	Separation
0.00	6.82	68×
0.05	6.91	71×
0.10	6.91	44×
0.15	6.86	343×
0.20	6.90	1,038×
0.30	7.21	3,110×
0.50	7.61	235×

Key findings:

1. **Manifold expansion**: D_{eff} increases from 6.82 (uncoupled) to 7.61 (strong coupling), a 12% expansion. This confirms the manifold expansion theorem [32]: coupling opens new dimensions in the accessible statistical manifold.

2. **Separation peaks at intermediate coupling**: The separation ratio maximizes around $\kappa \approx 0.30$ (3,110×), then collapses at high coupling. This reflects the transition from coordination to synchronization: moderate coupling enables coordination equilibria; excessive coupling forces all compartments into the same state.

3. **Nonlinear regime dependence**: The non-monotonic separation curve demonstrates that code emergence is not simply "more coupling = better" but requires tuning to the intermediate-coupling regime where compartments influence but do not dominate each other.

This validates the theoretical claim: the simulation is not merely demonstrating code emergence but *quantitatively testing the manifold expansion prediction* from information geometry.

4.11 Ablation Studies

We validate that code emergence depends on the claimed mechanisms through systematic ablations:

Condition	Unique Codes	Separation	Bimodality
Full system	32/32	335,000×	89% saturated
Channel blocked (noise input)	8/32	0.6×	34% saturated
No substrate competition	3/32	2×	22% saturated
Random spatial projections (5/5 work)	24–32/32	29–8,800×	89% saturated
No clipping (pre-clip values)	32/32	335,000×	87% saturated
No mean-centering	28/32	180,000×	82% saturated

Key findings:

1. **Channel matters**: Blocking the code signal (replacing with noise) destroys separation (0.6×), confirming information flows through the channel.

2. **Substrate competition is essential**: Removing competition collapses unique codes from 32 to 3 and bimodality from 89% to 22%.

3. **Discretization is intrinsic**: The no-clipping test shows bimodality persists without numerical thresholding—discretization comes from chemistry, not artifacts.

4. **Readout is robust**: Random electrode placements (5/5 tested) still yield separable codes, demonstrating discreteness is a field property, not a measurement artifact.

5. **Mean-centering is helpful but not essential**: Removing it reduces unique codes from 32 to 28 but the system remains functional.

4.12 Scale Dependence

To test whether code emergence persists at larger scales, we ran simulations at 169 vesicles (7-ring hexagonal array) with 512 internal dimensions per compartment:

Metric	Medium (61×128D)	Massive (169×512D)
Unique codes	32/32	32/32
D_{eff}	6.3	16.7
Env–Attractor correlation	0.72	0.83
Reproducibility	98.4%	84.2%
Separation ratio	335,000×	4.4×

Key findings:

1. **Code discrimination persists**: All 32 environmental configurations still map to 32 unique output sequences at massive scale.

2. **Higher expressivity**: D_{eff} increases from 6.3 to 16.7, and environment–attractor correlation improves from 0.72 to 0.83. The larger system uses more of its available dimensions and preserves environmental structure better.

3. **Stability–expressivity tradeoff**: Reproducibility drops from 98.4% to 84.2%, and separation ratio drops dramatically. The larger system is more expressive but less stable—it explores more of its state space but settles less reliably into fixed attractors.

This tradeoff is consistent with the theoretical framework: larger systems have more degrees of freedom to coordinate, enabling richer codes, but also more ways to drift between runs. The intermediate scale (61 vesicles, 128D) appears to be a "sweet spot" where the system has enough expressivity for reliable discrimination while remaining stable enough for reproducible codes. This may explain why early protocell networks would have operated at modest scales before selection pressure optimized for reliability.

5 Discussion

5.1 Abiogenesis: The "Between" Becoming "Within"

We propose that abiogenesis follows a three-phase trajectory:

1. **Phase 1 (The "Between")**: Coherence dynamics connect simple compartments via coupling fields. The "code" is volatile, living in the geometry of inter-compartment coordination—a distributed "ghost" in the electric/chemical fields between cells.

2. **Phase 2 (The Pressure)**: To survive environmental fluctuations and predict neighbor behavior, a compartment must internalize a compressed model of the distributed field dynamics.

3. **Phase 3 (The "Within")**: The compartment creates stable internal references (eventually RNA/DNA) to model the external field. This is the **Compression Transition**—the crystallization of "between" into "within."

This explains why internal symbols match external signals without a designer: the internal code is a *learned compression* of the coordination dynamics that already existed in the coupling fields. The genetic code is not the origin of meaning; it is the *fossilization* of meaning that lived in the distributed state.

5.2 The DNA-RNA-Protein System as Communication

On this view, the DNA-RNA-protein system is literally a communication system—not metaphorically:

- **DNA**: Persistent storage (slow, stable channel)

- **RNA**: Transmittable message (routing + control)

- **Protein**: Embodied actuation (receiver output)

But this is *late* communication—the internalization of external protocols. The original "messages" were the coordination patterns in the coupling fields between protocells.

5.3 Implications for Artificial Life

Our framework suggests a recipe for artificial life: don't try to "program" life. Create bounded, high-dimensional, far-from-equilibrium systems with weak coupling and multiscale dynamics. Codes will emerge.

The key ingredients are:

1. High-dimensional internal dynamics (many coupled degrees of freedom)

2. Weak coupling between compartments (influence without collapse)

3. Multiple signaling channels (redundancy and temporal separation)

4. Environmental heterogeneity (symmetry breaking)

5. Substrate competition (emergent discretization)

5.4 Scope and Limitations

We are explicit about what this paper demonstrates and what remains hypothetical:
Demonstrated in this paper:

- A minimal coupled-compartment dynamical system with resource competition yields discrete, reproducible, decodable boundary patterns

- Discretization emerges from mass-action kinetics (substrate competition), not external logic

- The mechanism is robust across random initializations (20/20 seeds produce functional codes)

- Ablations confirm the causal role of coupling and competition

Limitations:

- **Phenomenological reservoir**: The internal dynamics use a generic high-dimensional system (reservoir), not a literal reaction network. We do not claim a specific stoichiometry.

- **External observer for decoding**: Decoding accuracy is measured by an external observer comparing attractor states. A fully autonomous decoder would require additional "readout chemistry" inside the receiver.

- **Discrete environment labels**: The current simulations use 5-bit environmental configurations. The key continuity is in internal trajectories; demonstrating discretization of truly continuous environmental gradients remains future work.

- **No evolutionary dynamics**: We show code emergence in a fixed architecture. The evolutionary pathway from coupled dynamics to compressed genetic codes is hypothesized, not simulated.

Hypothesis/extrapolation: Early protocell networks could have exploited such coordination regimes; genetic encoding later compresses and internalizes them. This is a plausible scenario, not a demonstrated fact.

5.5 Mathematical Foundations

The simulation results presented here are underpinned by two formal information-geometric results:

1. Manifold expansion under coupling [32]. When dynamical systems are coupled, the Fisher information geometry of the accessible family can expand—the Fisher rank of the joint system may exceed the sum of individual ranks. This "superadditive complexity" provides the theoretical basis for why coupled compartments can generate richer coordination patterns than isolated compartments. The manifold expansion theorem proves that coupling can open new parametric directions in the statistical manifold, explaining how discrete codes emerge from continuous dynamics.

2. The observable dimensionality bound [33]. High-dimensional systems observed through low-capacity channels create fundamental limits on external inference. When internal dynamics have effective dimensionality D_{eff} exceeding the observer's channel capacity C_{obs}, the system exhibits "curvature amplification"—small perturbations in internal state produce exponentially separated geodesics in the observable projection. This explains why coupled protocell networks can maintain stable, reproducible codes: the high-dimensional internal dynamics create attractor basins that are robust to perturbation while remaining distinguishable at the boundary.

Together, these results establish that code emergence is not a special property of our particular simulation but a *geometric consequence* of high-dimensional coupling observed through low-capacity interfaces. The simulations demonstrate this mechanism; the mathematics proves its generality.

5.6 Experimental Accessibility

A practical advantage of the vesicle model is experimental tractability. Modern microfluidic techniques allow:

- Fabrication of vesicle arrays with controlled topology [20]

- Loading with oscillatory chemistry (e.g., BZ reaction, glycolytic oscillators)

- Measurement of inter-vesicle coupling via fluorescence, pH indicators, or voltage-sensitive dyes

- Systematic variation of coupling strength, membrane composition, and network geometry

The predictions of our model—that coupled vesicle networks will spontaneously develop reproducible, discrete, decodable boundary patterns—are directly testable with existing laboratory techniques.

6 Conclusion

We have argued that symbolic codes first emerged as coordination interfaces between coupled protocellular compartments, not as transmitted messages requiring pre-existing machinery. The key mechanism is substrate competition, which discretizes continuous dynamics into stable symbols without external logic.

Our simulations demonstrate that networks of coupled compartments with high-dimensional internal dynamics reliably produce:

- 32 unique symbol sequences from 32 environmental configurations (for a fixed network)

- Near-perfect reproducibility across trials (mean 98.4% across 20 random seeds)

- High decoder accuracy (mean 99%) via receiver attractor separation

- Emergent discretization via mass-action kinetics (89% bimodality)

The deeper claim is that **communication precedes information storage**. Genes are not the origin of codes; they are the compression of codes that already existed as coordination equilibria in protocellular networks. Abiogenesis is the history of the "between" becoming the "within"—life first exists in the distributed coupling fields, then internalizes those dynamics as genetic memory.

Life runs on coordination. The code is not a message. The code is a convention. And the genetic code is the fossilization of meaning that once lived in the electric fields between cells.

Code availability: Simulation code is available at `https://github.com/todd866/protocell-codes`
Companion paper: For a formal information-geometric treatment of manifold expansion under high-dimensional coupling, see: I. Todd, "Communication Beyond Information: Manifold Expansion via High-Dimensional Coupling," available at `https://github.com/todd866/manifold-expansion`

Acknowledgments

References

[1] M. Levin. Bioelectric signaling: Reprogrammable circuits underlying embryogenesis, regeneration, and cancer. *Cell*, 184(6):1971–1989, 2021.

[2] M. Levin. Molecular bioelectricity in developmental biology: New tools and recent discoveries. *BioEssays*, 34(3):205–217, 2012.

[3] D. S. Adams and M. Levin. Endogenous voltage gradients as mediators of cell-cell communication: strategies for investigating bioelectrical signals during pattern formation. *Cell and Tissue Research*, 352(1):95–122, 2013.

[4] R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[5] M. A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, 2006.

[6] B. Skyrms. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, 2004.

[7] J. W. Szostak, D. P. Bartel, and P. L. Luisi. Synthesizing life. *Nature*, 409(6818):387–390, 2001.

[8] I. A. Chen and P. Walde. From self-assembled vesicles to protocells. *Cold Spring Harbor Perspectives in Biology*, 2(7):a002170, 2010.

[9] P. Adamski, M. Eleveld, A. Sood, A. Kun, A. Szilágyi, T. Czárán, E. Szathmáry, and S. Otto. From self-replication to replicator systems *en route* to de novo life. *Nature Reviews Chemistry*, 4(8):386–403, 2020.

[10] N. Lane and W. Martin. The energetics of genome complexity. *Nature*, 467(7318):929–934, 2010.

[11] D. W. Deamer and J. P. Dworkin. Chemistry and physics of primitive membranes. *Topics in Current Chemistry*, 259:1–27, 2005.

[12] J. P. Dworkin, D. W. Deamer, S. A. Sandford, and L. J. Allamandola. Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices. *Proceedings of the National Academy of Sciences*, 98(3):815–819, 2001.

[13] T. M. McCollom, G. Ritter, and B. R. T. Simoneit. Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Origins of Life and Evolution of the Biosphere*, 29:153–166, 1999.

[14] A. I. Rushdi and B. R. T. Simoneit. Lipid formation by aqueous Fischer-Tropsch-type synthesis over a temperature range of 100 to 400°C. *Origins of Life and Evolution of the Biosphere*, 31:103–118, 2001.

[15] W. Martin and M. J. Russell. On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B*, 362:1887–1926, 2007.

[16] S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco, and J. W. Szostak. Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, 454:122–125, 2008.

[17] I. Budin and J. W. Szostak. Expanding roles for diverse physical phenomena during the origin of life. *Annual Review of Biophysics*, 39:245–263, 2010.

[18] T. F. Zhu and J. W. Szostak. Coupled growth and division of model protocell membranes. *Journal of the American Chemical Society*, 131:5705–5713, 2009.

[19] I. Budin and J. W. Szostak. Physical effects underlying the transition from primitive to modern cell membranes. *Proceedings of the National Academy of Sciences*, 108:5249–5254, 2011.

[20] M. Karlsson, K. Sott, M. Davidson, A.-S. Cans, P. Linderholm, D. Chiu, and O. Orwar. Formation of geometrically complex lipid nanotube-vesicle networks of higher-order topologies. *Proceedings of the National Academy of Sciences*, 99(18):11573–11578, 2002.

[21] G. Bolognesi, M. S. Friddin, A. Salehi-Reyhani, N. E. Sherlock, O. Ces, and Y. Sherlock. Sculpting and fusing biomimetic vesicle networks using optical tweezers. *Nature Communications*, 9:1882, 2018.

[22] J. Maynard Smith and E. Szathmáry. *The Major Transitions in Evolution*. Oxford University Press, 1995.

[23] R. E. Michod. *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton University Press, 1999.

[24] P. Turchin. *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth*. Beresta Books, 2016.

[25] S. H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, 143(1–4):1–20, 2000.

[26] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 2001.

[27] K. Friston. A free energy principle for a particular physics. *arXiv preprint arXiv:1906.10184*, 2019.

[28] J. L. England. Statistical physics of self-replication. *Journal of Chemical Physics*, 139(12):121923, 2013.

[29] I. Todd. The limits of falsifiability: Dimensionality, measurement thresholds, and the sub-Landauer domain in biological systems. *BioSystems*, 258:105608, 2025.

[30] I. Todd. Timing inaccessibility and the projection bound: Resolving Maxwell's demon for continuous biological substrates. *BioSystems*, 258:105632, 2025.

[31] I. Todd. Intelligence as high-dimensional coherence: The observable dimensionality bound and computational tractability. *BioSystems*, under review, 2026. Preprint: `https://github.com/todd866/intelligence-biosystems`

[32] I. Todd. Communication beyond information: Manifold expansion via high-dimensional coupling. In preparation for *Information Geometry*, 2026. Preprint: `https://github.com/todd866/manifold-expansion`

[33] I. Todd. Curvature amplification in high-to-low dimensional projections: Limits on external inference. In preparation for *Information Geometry*, 2026. Preprint: `https://github.com/todd866/tracking-complexity`

[34] S. Bhattacharya and S. Bhattacharya. Multi-modal signaling in development: A paradigm beyond the genetic code. *Current Opinion in Systems Biology*, 11:82–90, 2018.

[35] B. Perbal. Communication is the key. *Cell Communication and Signaling*, 1(1):1–4, 2003.

[36] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Wiley-Blackwell, 4th edition, 2012.

[37] C. C. Grueter, B. Chapais, and D. Zinner. Evolution of multilevel social systems in nonhuman primates and humans. *International Journal of Primatology*, 33(5):1002–1037, 2012.