Codes as Coordination: A Physical System that Generates Digital Communication Without Pre-Programming

Ian Todd

Sydney Medical School, University of Sydney

`itod2305@uni.sydney.edu.au`

January 2026

1 Executive Summary

We present a purely chemical architecture that generates, transmits, and decodes digital symbol sequences without pre-programmed mapping.

Requirement	Our System	Status
Encoder	Network of 61 coupled compartments	✓
Message	4-symbol sequence per configuration	✓
Decoder	Physics-based receiver colony	✓
≥ 32 states	32 distinguishable states (100% decoder accuracy)	✓
Two-layer ($n + k \geq 5$)	4 symbols $\times$ 10 bits $= n = 4$, $k = 10$	✓
Digital	Emergent bi-stability via mass-action kinetics	✓
No pre-programming	Environmental heterogeneity, not logic	✓
No biological material	Synthetic chemistry only	✓

Core mechanism: Codes emerge as coordination equilibria between coupled protocellular compartments. Discretization arises from substrate competition (lateral inhibition), not engineered logic.

2 System Architecture

2.1 Components

Component	Function	Implementation
Compartments	Semi-autonomous agents	61 vesicles in hexagonal array
Internal dynamics	High-dimensional chemistry	128 dimensions per compartment
Discretization	Emergent symbol formation	Substrate competition (30 channels)
Coupling	Neighbor communication	Weak boundary signal exchange
Spatial structure	Symmetry breaking	Center-edge, gradient differentiation

2.2 How It Works

Each compartment contains high-dimensional nonlinear reaction dynamics (128 coupled species near the edge of instability). Output channels compete for finite substrate via **mass-action kinetics**:

- **Hill kinetics**: Saturation function $S^n/(K^n + S^n)$ with cooperativity $n = 2$

- **Substrate competition**: Allocation $\propto \text{activity}^n / \sum(\text{activity}^n)$

- This is the **quasi-steady-state (QSSA)** solution to competitive binding—standard biochemistry (Michaelis-Menten, competitive inhibition), not engineered logic

Compartments are coupled through boundary signals: each vesicle's state is influenced by the average readout of its neighbors. This creates coordination pressure without global mixing.

Environmental heterogeneity drives differentiation: vesicles at different spatial locations experience different stimulus conditions (center vs. edge, top vs. bottom). This breaks degeneracy between input configurations. *Crucially, these gradients represent non-informational geometric constraints*—e.g., a rock shading part of a tide pool, proximity to a heat source, or 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.

Temporal forcing creates genuine sequence structure: each of 4 cycles experiences different environmental conditions (modeling diurnal variation, tidal rhythms).

3 Results

3.1 Performance Summary

Metric	Result
Compartments	61 vesicles (hexagonal array)
Internal dimensions	128 per compartment
Readout channels	30
Unique symbol sequences	24/32 (8 collisions at symbol level)
Encoder reproducibility	100%
Separation ratio (between/within)	243×
Bimodal fraction ($|x| > 0.5$)	89%
Decoder accuracy	100% (all 32 distinguishable)

Key results:

- **100% decoder accuracy**: physics-based receiver distinguishes all 32 inputs

- **100% encoder reproducibility**: same input $\rightarrow$ same output across trials

- **24 unique symbol sequences**: discretization loses some information, but the full 20D transmitted signal remains distinguishable

- **Emergent discretization**: 89% of readout values are saturated ($|x| > 0.5$)

3.2 Encoding Table (Full Codebook)

Complete 4-symbol character sequences for all 32 configurations:

Config	Binary	S_1	S_2	S_3	S_4	Note
0	00000	0100001000	0100001000	0100001000	0100001000	unique
1	00001	0000000000	0000000100	0000000100	0000000000	unique
2	00010	0000001000	0000001000	0000001000	0000000000	unique
3	00011	0000000000	0000000000	0000100000	0000000000	unique
4	00100	0000000000	0000000000	0000000000	0000100000	unique
5	00101	0000000000	0000000000	0000000000	0000000000	†
6	00110	0000000010	0000000010	0000000010	0000000010	unique
7	00111	0000100000	0000100000	0000100000	0000100000	= 22
8	01000	0100001000	0100001000	0100001000	0100001000	= 0
9	01001	0000000000	0000000000	0000000000	0000000000	†
10	01010	0100001000	0100001000	0100001000	0100000000	unique
11	01011	0000000000	0010000000	0010000000	0010000000	unique
12	01100	0100000000	0100000000	0100000000	0100000000	unique
13	01101	0000000000	0000000100	0000000000	0000000000	unique
14	01110	0000000000	0000100000	0000100000	0000000000	unique
15	01111	0000000010	0000000000	0000000000	0000000000	unique
16	10000	0000001000	0000001000	0000001000	0000001000	= 24
17	10001	0000001100	0000001100	0000001100	0000001100	unique
18	10010	0000001000	0000001000	0010001000	0010001000	unique
19	10011	0010000000	0010000000	0010000000	0010000000	unique
20	10100	0000001000	0000001000	0000000000	0000000000	unique
21	10101	0000000000	0000000000	0000000000	0000000000	†
22	10110	0000100000	0000100000	0000100000	0000100000	= 7
23	10111	0000000010	0000000000	0000000000	0000000010	unique
24	11000	0000001000	0000001000	0000001000	0000001000	= 16
25	11001	0000000100	0000000100	0100000100	0000000100	unique
26	11010	0000001000	0000001000	0000001000	0000011000	unique
27	11011	0000000000	0000000000	0000000000	0000000000	†
28	11100	0000000000	0000000000	0000000000	0000000000	†
29	11101	0000000100	0000000100	0000000100	0000000100	unique
30	11110	0000000000	0000010000	0000010000	0000010000	unique
31	11111	0000000010	0000000010	0000000010	0000000010	unique

† = maps to null sequence (collision); = N = identical to config N. 24 unique sequences, 8 symbol-level collisions. Despite collisions at the symbol level, the decoder achieves 100% accuracy using the full 20D transmitted signal.

Two-layer structure: Each character is a 4-symbol sequence. Each symbol is 10 bits (from center readout channels). Thus $n = 4$ symbols, $k = 10$ bits per symbol, satisfying $n + k = 14 \geq 5$.
Note on signal dimensionality: The transmitted physical signal is 20D (center + edge aggregates), but each **symbol** is defined as the 10-bit sign pattern of the center channels. Edge channels provide redundancy and enable 100% decoder accuracy despite symbol-level collisions.

3.3 Physics Decoder

The decoder is a **second vesicle array** (the "receiver colony") with the same dynamics but *different random internal structure*:

- Receives encoder's 20-dimensional output signal (center + edge aggregates)

- Processes through its own nonlinear reaction dynamics (independent random wiring)

- Output pattern compared to canonical receiver responses (not encoder patterns)

- **No machine learning, no training**—pure physics

Crucially, the receiver has different random internal chemistry than the encoder. This proves the code is robust to specific internal wiring—the information is in the interface, not the substrate.

3.3.1 Decoding Rule (Objective and Determinable)

For each configuration c, define the **canonical receiver response $\mathbf{R}_c = (r_{c,1}, \ldots, r_{c,4})$** as the mean receiver emission over $N \geq 10$ calibration trials.

For a new received message $\mathbf{r} = (r_1, \ldots, r_4)$, classification is:

$$c^* = \arg\min_c \sum_{t=1}^{4} \|\mathbf{r}_t - \mathbf{R}_{c,t}\|^2$$

Verification: This rule achieves 100% accuracy (32/32 correct) across all test trials. The confusion matrix is diagonal-dominant with no off-diagonal entries exceeding 1%.
Key point: The receiver colony is "blind" to the environment—it sees *only* the transmitted signal. If it correctly reconstructs the input configuration, the information must be in the code, not the environment.

4 Verification Protocol

4.1 Discretization Test (Proving "Digital")

- **Saturation ratio**: $\geq 85\%$ of all output states must fall within saturated basins ($|x| > 0.5$).
  Achieved: 89%

- **Bimodality coefficient**: Distribution of boundary signals must yield Sarle's BC > 0.555, with histogram showing two distinct peaks separated by a density valley (the "forbidden analog zone")

4.2 Reproducibility Test (Proving "Stable")

- **Intra-class stability**: Average Hamming distance between symbol sequences from the *same* input across $N \geq 10$ trials must be $< 5\%$ of sequence length

- **Cohen's Kappa**: $\kappa > 0.8$ for symbol identity across repeats, indicating "almost perfect" agreement beyond chance. *Achieved: $\kappa = 1.0$*

4.3 Distinguishability Test (Proving "Information")

- **Separation ratio**: Between-cluster variance / within-cluster variance $> 100\times$. *Achieved: 243×*

- **Unique symbol sequences**: 24/32 at symbol level (8 collisions); but 32/32 distinguishable via full transmitted signal

- **Decoder accuracy**: Blind physics-based receiver achieves $\geq 95\%$ classification. *Achieved: 100%*

4.4 Null-Model Controls (Proving "Emergence")

- **Dead chemistry**: Without oscillatory dynamics, output correlates with input intensity but lacks bimodality and sequence structure

- **Scrambled topology**: Randomly rewired couplings cause reproducibility to drop below 50% and separation ratio to collapse below $10\times$

- **Channel blockade**: Blocking boundary signaling eliminates emergent symbols (collapse to trivial uniform state)

- **All reagents synthetic**: No biological material; encoding table is discovered, not designed

4.5 Ablation Results Summary

Condition	Unique Codes	Separation	Bimodality
Full system	24/32	$243\times$	89% saturated
Channel blocked	8/32	$0.6\times$	34% saturated
No substrate competition	3/32	$2\times$	22% saturated
Random projections	24/32	29–$8862\times$	89% saturated
No clipping (numerical)	24/32	$243\times$	89% saturated

Key findings: (1) Digitality depends on substrate competition, not numerical artifacts—the "no clipping" test confirms bimodality persists without thresholding. (2) Emergent codes are a property of the *field dynamics*, not electrode placement—random spatial projections also yield separable codes (5/5 success).

5 Physical Implementation

The architecture maps to laboratory-realizable systems:

Component	Physical Realization
Compartments	Lipid vesicles or microfluidic droplets in array
Internal dynamics	BZ-type oscillatory chemistry
Coupling	Diffusion through shared medium + membrane contact
Boundary readout	pH-sensitive dye + voltage-sensitive indicators
Environmental forcing	UV exposure, temperature gradients, ion fluxes

5.0.1 Measurement Apparatus

The readout uses **differential measurement**: each channel's signal is recorded relative to the mean field (common-mode subtraction). This is standard electrochemistry practice—e.g., electrode potentials measured against a reference electrode.

Physically, this corresponds to:

- Reporter species in redox equilibrium (signal = deviation from equilibrium potential)

- Differential dye systems (e.g., ratiometric pH indicators)

- Common-mode rejection in optical readout

The "mean-centering" in our simulation models this physical measurement reference frame—it is part of the readout hardware, not computational logic.

On the gain factor: The amplification in our simulation represents the sensitivity of the physical measurement apparatus (e.g., voltage-sensitive dye quantum yield, electrode gain). The bimodality exists in the chemical allocation ratios; the gain merely makes it observable. Crucially, the "no-clip" validation confirms bimodality persists without any numerical thresholding.

Experimental protocol:

1. Prepare hexagonal vesicle array with controllable coupling

2. Load with redox-active, pH-buffered oscillatory reaction mixture

3. Add boundary indicators (encapsulated pH dye, precipitation system)

4. Apply 32 forcing configurations (5-bit environmental input)

5. Record boundary states over 4 temporal cycles per configuration

6. Repeat 10 trials per configuration for reproducibility statistics

7. Feed encoder output to receiver colony and record response

6 Why This Works: The Physics of Discretization

The mechanism responsible for discretizing continuous internal dynamics into binary symbols is not an engineered logic gate, but a direct consequence of **mass-action kinetics** in a resource-constrained system.

6.1 Substrate Competition via QSSA

Multiple output channels ($i = 1, \ldots, n$) compete for a finite, shared substrate pool (S_{total}). Following standard competitive binding kinetics [1], the fractional allocation to channel i under the Quasi-Steady-State Assumption is:

$$\text{Allocation}_i = \frac{(a_i)^h}{\sum_j (a_j)^h}$$

where a_i is the activity of channel i and h is the Hill coefficient representing allosteric cooperativity.

This equation, which *mathematically resembles* the "Softmax" function used in machine learning, here arises from **conservation of mass**. The "sum" in the denominator is not a calculated normalization—it is the physical reality that a substrate molecule consumed by Channel A is unavailable to Channel B.

6.2 Why This Produces Digital Output

- **Winner-take-most dynamics**: When $h > 1$, small differences in activity are amplified into large differences in allocation

- **Bimodality**: The system spends 89% of time in saturated states ($|x| > 0.5$), with minimal occupancy in the "analog" transition zone

- **No threshold engineering**: Discretization emerges from enzyme kinetics, not programmed logic gates

Operational definition of "digital": We define a bit by the sign of the readout channel after equilibration. The distribution is bimodal with a low-occupancy transition region, so the readout is digital under this operational definition.

The coupling between compartments then allows these local discretizations to coordinate into global patterns—the "code" emerges as a coordination equilibrium.

References

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

7   Theoretical Foundation

The mechanism demonstrated here rests on two mathematical results:

1. **Manifold expansion**: When high-dimensional systems couple, the identifiable parameter space can grow superadditively—the coupled system has more distinguishable states than the sum of its parts. This explains why 61 coupled compartments generate 32 unique codes that no single compartment could produce alone.

   Mathematical treatment: "Communication Beyond Information: Manifold Expansion via High-Dimensional Coupling," available at `https://github.com/todd866/manifold-expansion`

2. **Observable dimensionality bound**: When internal dynamics exceed observer channel capacity, stable codes *must* form as coordination equilibria. The discretization we observe is not accidental—it is thermodynamically necessary when high-D dynamics collapse to low-D outputs.

   Mathematical treatment: "Curvature Amplification of Tracking Complexity on Statistical Manifolds," available at `https://github.com/todd866/tracking-complexity`

Together: **coupling creates complexity that is self-protecting**. The codes emerge because coupling expands the accessible manifold; they stabilize because high-dimensional dynamics exceed what observers can track.

8   Conclusion

We demonstrate a physical system where:

1. An encoder (coupled compartment network) maps 32 environmental configurations to 24 unique 4-symbol sequences (some collisions at the symbol level, but 32 distinguishable via full signal)

2. A decoder (physics-based receiver colony) distinguishes all 32 with 100% accuracy

3. Discretization emerges from mass-action kinetics, not pre-programmed logic

4. The encoding table is discovered, not designed

This satisfies all prize requirements. The system is experimentally realizable with existing microfluidic and oscillatory chemistry techniques.

Code availability: `https://github.com/todd866/protocell-codes`
Contact: Ian Todd, `itod2305@uni.sydney.edu.au`