The Origin of Codes: How Digital Information Emerges from High-Dimensional Chemistry

Ian Todd
Sydney Medical School
University of Sydney
Sydney, NSW, Australia
itod2305@uni.sydney.edu.au

Abstract

The Evolution 2.0 challenge asks how coded information arises from non-living chemistry. We argue the question has the causality backwards. Digital codes are not *built into* chemistry; they *emerge from* the projection of high-dimensional dynamics onto low-dimensional readouts. When diverse chemical species track a common external forcing (the sun, tides), their collective response forms a high-dimensional attractor—many "ways to track the sun." At boundaries (membranes, surfaces), this high-D dynamics projects onto discrete symbols. The transition to life occurs when compartmentalization (lipid bilayers) allows chemistry to swap the external attractor for an endogenously generated one—self-organization replacing sun-organization. We do not need to engineer codes; we need to understand why they inevitably emerge.

1 The Challenge Has It Backwards

The Evolution 2.0 prize asks: *How does coded information (encoder → message → decoder) arise from purely non-living chemistry?*

This framing assumes:

- Digital codes are fundamental and must be *built*
- We need to engineer an encoder that produces symbols
- The chemistry must somehow "generate" discrete states

We argue this is backwards. The actual story is:

- High-dimensional continuous dynamics are fundamental
- "Digital" emerges when high-D projects onto low-D at a boundary

- Codes are not designed—they are *discovered*
- The discreteness is in the *readout*, not the chemistry

Core claim: Any sufficiently diverse chemistry, tracking a common external forcing, read out at a boundary, automatically produces codes. You don't build codes into chemistry. You find them at the interface.

2 High-Dimensional Attractors from Common Forcing

2.1 The Setup

Consider prebiotic Earth:

- **Diverse chemistry**: Amino acids, sugars, nucleobases, lipids, metal ions, thiols, phosphates—hundreds of species from atmospheric synthesis, meteoric input, hydrothermal vents
- **Common forcing**: The sun (day/night UV, temperature cycling), tides (wet/dry, concentration/dilution), seasons
- **Cross-reactivity**: Everything can react with everything—condensation, complexation, redox, catalysis

Each chemical species responds to the forcing in its own way:

- Amino acids: photodegradation during day, condensation during dry phases
- Sugars: isomerization rates temperature-dependent, formose branching UV-sensitive
- Nucleobases: UV absorption drives chemistry, metal complexation shifts with redox
- Lipids: phase transitions with temperature, self-assembly concentration-dependent

2.2 The High-Dimensional Attractor

All these species must track the *same* external rhythm. But each tracks it differently.

The collective response is high-dimensional because:

- Many species ($n \gg 1$) with partially independent dynamics
- Cross-coupling prevents lockstep motion
- The "attractor" is not a single state but a *manifold* of ways to track the forcing

Different initial conditions lead to different positions on this attractor manifold. These are the proto-codes: distinguishable, reproducible, history-dependent states.

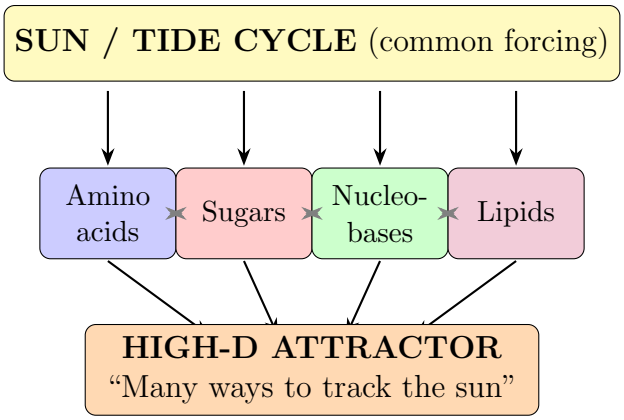

Figure 1: Diverse chemistry tracking a common forcing produces a high-dimensional attractor. Each species responds to the sun/tide cycle in its own way; their collective, coupled response spans many dimensions.

3 Digital Emerges at Boundaries

3.1 The Projection Principle

The chemistry itself is continuous and high-dimensional. Where do discrete "symbols" come from?

Answer: From projection onto low-dimensional readouts at boundaries.

Definition 1 (Emergent Digitality)**.** *A system exhibits emergent digitality when a high-dimensional continuous dynamics, projected onto a low-dimensional boundary observable, produces distinguishable, reproducible, discrete states—without those states being built into the dynamics.*

3.2 What Boundaries Do

Boundaries—membranes, mineral surfaces, phase interfaces—are selective:

- **Membranes**: Only certain molecules permeate; the flux depends on the full chemical state but the *output* is which molecules made it through
- **Surfaces**: Adsorption is selective; different attractor states produce different adsorption patterns
- **Phase interfaces**: Partitioning depends on hydrophobicity, charge, etc.; the oil/water distribution reflects the interior state

The boundary acts as a **coarse-graining operator**. It takes the high-D interior state and projects it onto a few observable dimensions.

3.3 Why This Creates Discrete Symbols

When a high-D attractor projects onto a low-D observable, the image is not uniform. It has:

- **Clusters**: Regions of the attractor that map to similar outputs
- **Gaps**: Outputs that no attractor state produces
- **Bistability**: Boundary mechanisms (pH indicators, precipitation thresholds) that snap to discrete states

The discreteness is not in the chemistry. It is in the *interface between chemistry and readout*.

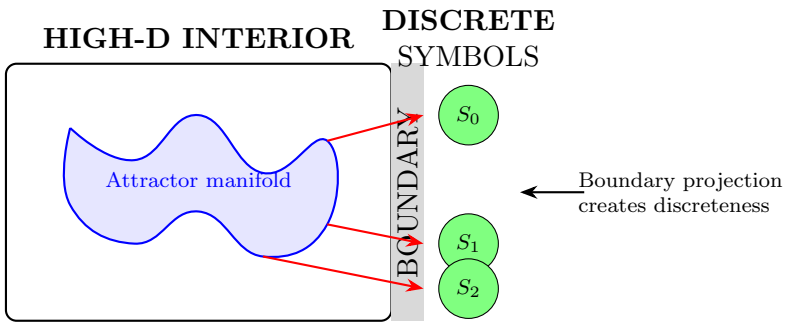

Figure 2: Continuous high-D dynamics become discrete symbols at boundaries. The attractor manifold (blue) projects through the boundary onto distinguishable outputs. Discreteness is not built into the chemistry; it emerges from the projection.

4 The Transition to Life

4.1 Stage 1: External Organization

Initially, the chemistry tracks an *external* attractor—the sun/tide cycle. The codes that emerge are:

- Determined by the external forcing
- Reproducible (same forcing $\rightarrow$ same attractor region)
- History-dependent (initial conditions matter)

But the chemistry is not yet "alive." It is organized *by* the sun, not *by itself*.

4.2 Stage 2: Compartmentalization

Lipid bilayers form spontaneously in prebiotic conditions. When they encapsulate a subset of the chemistry:

- The interior becomes *concentrated*
- The boundary becomes *defined* (the membrane)
- The interior chemistry can develop its own dynamics, partially decoupled from the external forcing

4.3 Stage 3: Swap the Attractor

The key transition: inside the compartment, some chemistry becomes **autocatalytic**—self-amplifying reaction networks that can sustain themselves.

Now the interior dynamics can generate their *own* attractor, independent of the sun.

> **The transition to life:** Swap the external attractor (sun-tracking) for an endogenously generated attractor (self-organization). Package it in a lipid bilayer. The "code" is now self-maintained.

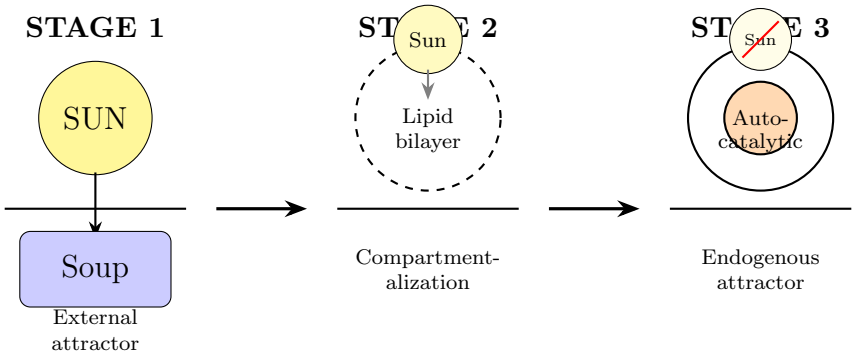

Figure 3: The transition to life. Stage 1: Chemistry tracks external attractor (sun). Stage 2: Compartmentalization in lipid bilayer. Stage 3: Internal autocatalytic network generates its own attractor; external forcing becomes optional. Life = chemistry that organizes itself.

5 What "Digital" Actually Means

The Evolution 2.0 challenge demands a "digital" system. But what does digital mean in chemistry?

5.1 The Wrong Answer

Wrong: The chemistry has discrete states built in.

This is the assumption behind looking for "chemical bits"—molecules that are either ON or OFF, reactions that are either HAPPENED or DIDN'T.

The problem: chemistry is fundamentally continuous. Concentrations vary smoothly. Reaction rates are continuous functions of temperature, pH, etc.

5.2 The Right Answer

Right: The readout has discrete states, because the boundary mechanism is bistable/multistable.

"Digital" is not a property of the interior dynamics. It is a property of the *interface*.

Examples of physically discrete readouts:

- **pH indicators**: Sharp color transitions at threshold pH
- **Precipitation**: Either precipitate forms or it doesn't
- **Phase transitions**: Lipid gel vs fluid phase
- **Membrane permeation**: Above threshold concentration, molecules cross; below, they don't

In each case, the interior chemistry is continuous; the readout snaps to discrete values because of the *physics of the boundary*, not because of "digital chemistry."

5.3 Implications for Evolution 2.0

The challenge asks: "Show chemistry generating digital information."

Our answer: Chemistry doesn't generate digital information. Chemistry is continuous and high-dimensional. Digital information *emerges at boundaries* when high-D dynamics project onto bistable readouts.

If you want to see codes:

1. Take diverse chemistry (many species)
2. Apply common forcing (sun/tide cycle)
3. Read out at a boundary (membrane, surface)
4. The codes are there—you discover them, you don't design them

6 The Experimental Realization

6.1 System Design

Based on the above framework, a compliant experimental system is:

Substrate (the "soup"):

- Amino acids: Gly, Ala, Asp, Glu, Ser, Val, Leu, Ile, Pro, Phe (10 species)
- Sugars: Ribose, glucose, glyceraldehyde, formaldehyde, glycolaldehyde (5 species)
- Nucleobases: Adenine, guanine, cytosine, uracil (4 species)
- Metal ions: Fe^{2+}, Mg^{2+}, Zn^{2+}, Ca^{2+} (4 species)
- Phosphate, thiols (cysteine), fatty acids (2–3 species)
- Total: ∼25–30 primary species, plus reaction products

Forcing (the "sun"):

- UV cycling: 12h on / 12h off (or accelerated: 1h cycles)
- Temperature cycling: 20°C → 60°C → 20°C
- Wet/dry cycling: Partial evaporation then rehydration

Boundary (the "readout"):

- Option A: Mineral surface (montmorillonite clay)—measure adsorption pattern
- Option B: Lipid vesicles—measure permeation/internal composition
- Option C: Phase interface—measure partitioning into oil phase

Readout mechanism (bistable for digitality):

- pH indicator dyes in readout compartment (discrete color states)
- Precipitation assay (Ca^{2+} + oxalate: precipitate or not)
- Turbidity threshold (lipid aggregation: clear or cloudy)

6.2 Protocol

1. Prepare substrate mixture in buffer with mineral surface or vesicles
2. Define input conditions: 64 combinations of (forcing phase, initial pH, metal ratio, concentration)
3. Run system for 10–100 forcing cycles
4. Sample boundary readout at end of each cycle
5. Record discrete state (which indicator state, which precipitation pattern)
6. Repeat 10 trials per input condition

6.3 Expected Outcomes

- Different input conditions → different boundary patterns
- Patterns cluster into discrete states (symbols) without imposed thresholds
- Reproducibility >70% (same input → same symbol)
- ≥32 distinct symbols observed

The codes are *discovered* by clustering the observed outputs, not *designed* by choosing thresholds.

7 Mapping to Evolution 2.0 Requirements

Requirement	**Our System**	**Note**
Encoder	High-D soup + forcing	Attractor selection
Message	Boundary flux/composition	Chemical intermediates
Decoder	Bistable indicator	Physical discretization
Symbol	Discrete readout state	Emergent, not imposed
Character	Sequence over cycles	M^n states
≥32 states	$4^3 = 64$	Satisfies requirement
Digital	Boundary bistability	Physics, not binning
No pre-programming	Codes discovered	Tables populated empirically
No biological material	Synthetic reagents	All commercial

8 Why the Challenge is Backwards

The Evolution 2.0 challenge implicitly assumes:

1. Codes are designed/engineered
2. Digital is fundamental
3. We need to find "the right chemistry"

We have argued:

1. Codes emerge inevitably from high-D dynamics + boundaries
2. Digital is a projection artifact, not a fundamental property
3. Any sufficiently diverse chemistry works—there is no "right" chemistry

The transition to life was not the *invention* of codes. It was:

1. High-D chemistry tracking external forcing (codes emerge at boundaries)

2. Compartmentalization (lipid bilayer defines the boundary)

3. Swapping external attractor for endogenous one (self-organization)

4. Learning to read and replicate the boundary states (evolution)

Life didn't create information. It *exapted* the codes that physics was already producing at the interface between high-dimensional chemistry and low-dimensional readouts.

9 Conclusion

We have presented a framework for understanding how coded information emerges from non-living chemistry:

1. **High-dimensional dynamics**: Diverse chemistry tracking common forcing creates a high-D attractor ("many ways to track the sun")

2. **Boundary projection**: Discrete symbols emerge when high-D projects onto low-D at boundaries (membranes, surfaces)

3. **Emergent digitality**: The discreteness is in the readout physics (bistability), not the chemistry

4. **Transition to life**: Compartmentalization + autocatalysis allows chemistry to swap external for endogenous attractors

The Evolution 2.0 challenge asks how to build codes into chemistry. We have shown that codes are not built—they emerge. The question is not "how do we engineer an encoder?" but "why do boundaries inevitably produce symbols?"

The answer: because projection from high-D to low-D, through a bistable interface, is a discretizing operation. Codes are not invented. They are discovered at the interface between complex dynamics and simple readouts.

This is not a limitation to be engineered around. It is the mechanism by which life arose.

Submitted to: HeroX Evolution 2.0 Prize