

Emergent Codes: How Digital Information Arises from High-Dimensional Prebiotic Chemistry

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

Abstract

The origin of genetic codes is often framed as a problem of engineering: how did chemistry come to *encode* information? We argue this framing inverts the actual causality. Digital codes are not built into chemistry; they emerge from the projection of high-dimensional dynamics onto low-dimensional boundaries. When diverse chemical species track a common external forcing (solar and tidal cycles), their collective response forms a high-dimensional attractor—many distinct “ways to track the sun.” At boundaries (membranes, mineral surfaces), this high-dimensional dynamics projects onto discrete symbols through bistable physical mechanisms. The transition to life occurs when compartmentalization allows chemistry to swap the external attractor for an endogenously generated one. We develop this framework mathematically, showing that emergent digitality is a generic consequence of dimensionality reduction at interfaces, and discuss implications for understanding the origin of biological information.

Keywords: Origin of life; genetic code; emergence; dimensionality; prebiotic chemistry; information

1 Introduction

The genetic code presents a puzzle: how did discrete, symbolic information arise from continuous chemical processes? Standard approaches seek specific reaction mechanisms—particular molecules that could serve as primordial “bits” (Eigen, 1971; Kauffman, 1993). This framing assumes that digitality must be engineered into chemistry. Recent work has emphasized the algorithmic and informational aspects of life’s origin (Walker and Davies, 2013; Schrödinger, 1944), yet the mechanism by which continuous chemistry produces discrete codes remains unclear.

We argue for an alternative view: digital codes are not fundamental but *emergent*. They arise inevitably when high-dimensional continuous dynamics project onto low-dimensional

observables at boundaries. The discreteness is not a property of the chemistry but of the interface between complex internal dynamics and simple readouts.

This perspective reframes the origin-of-codes problem. Rather than asking “what chemistry produces digital information?” we ask “why do boundaries inevitably discretize high-dimensional dynamics?” The answer involves well-understood mathematics: projection from high to low dimensions, through bistable mechanisms, is a discretizing operation.

The structure of the paper is as follows. Section 2 develops the concept of high-dimensional attractors from common forcing. Section 3 shows how digital symbols emerge at boundaries. Section 4 describes the transition from externally-organized to self-organized systems. Section 5 discusses experimental implications. Section 6 concludes with broader implications for understanding biological information.

2 High-Dimensional Attractors from Common Forcing

2.1 The Prebiotic Setup

Consider the chemical environment of early Earth (Ruiz-Mirazo et al., 2014; Pross, 2012):

- **Diverse chemistry:** Amino acids, sugars, nucleobases, lipids, metal ions, thiols, phosphates—hundreds of species from atmospheric synthesis, meteoric delivery, and hydrothermal processes (Miller, 1953; Chyba and Sagan, 1992)
- **Common forcing:** Solar radiation (day/night UV flux, temperature cycling), tidal cycles (wet/dry, concentration/dilution), seasonal variation
- **Cross-reactivity:** Extensive reaction networks—condensation, complexation, redox chemistry, catalysis

Each chemical species responds to the external forcing differently:

- Amino acids undergo photodegradation during daylight, condensation during dry phases
- Sugar isomerization rates depend on temperature; formose branching is UV-sensitive
- Nucleobase chemistry is driven by UV absorption; metal complexation shifts with redox state
- Lipid phase transitions track temperature; self-assembly depends on concentration

2.2 Formation of High-Dimensional Attractors

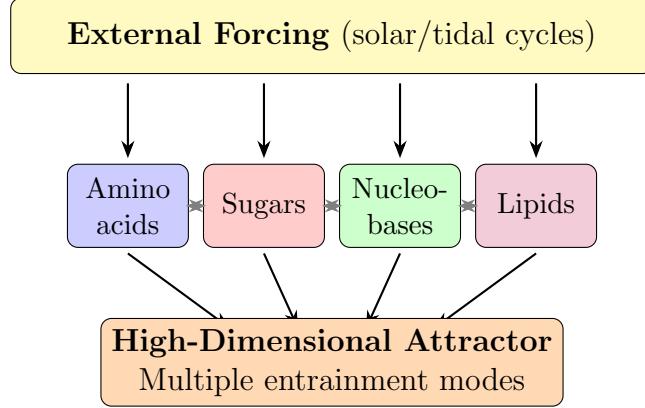
When multiple species with distinct response characteristics are coupled through cross-reactions, and all track a common external rhythm, the collective dynamics forms a high-dimensional attractor.

Let $\mathbf{x}(t) \in \mathbb{R}^n$ represent the concentrations of n chemical species. Under periodic forcing $F(t) = F(t + T)$, the dynamics satisfy:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})F(t) \quad (1)$$

where \mathbf{f} captures internal reaction dynamics and \mathbf{g} the forcing response.

For sufficiently diverse chemistry (large n , varied \mathbf{g}), the system admits multiple stable entrainment modes—different ways to synchronize with the external forcing. These modes form an attractor manifold $\mathcal{A} \subset \mathbb{R}^n$ with effective dimension $d_{\text{eff}} \ll n$ but still $d_{\text{eff}} \gg 1$.



Different initial conditions → different positions on attractor manifold

Figure 1: Diverse chemistry tracking a common forcing produces a high-dimensional attractor. Each species responds to the solar/tidal cycle distinctly; their coupled response spans multiple dimensions.

Definition 1 (Proto-code). *A **proto-code** is a distinguishable, reproducible, history-dependent position on the attractor manifold. Different initial conditions lead to different proto-codes, even under identical forcing.*

The key insight is that the external forcing does not determine a unique chemical state. It determines a *manifold* of states, parameterized by initial conditions and history. This multiplicity is the reservoir from which discrete codes will emerge.

3 Emergent Digitality at Boundaries

3.1 The Projection Principle

The chemistry itself is continuous. Discrete symbols arise from projection onto low-dimensional observables at boundaries.

Definition 2 (Emergent Digitality). *A system exhibits **emergent digitality** when high-dimensional continuous dynamics, projected onto a low-dimensional boundary observable, produces distinguishable, reproducible, discrete states—without those states being encoded in the dynamics.*

Let $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a projection operator representing the boundary readout, with $m \ll n$. The boundary observable is $\mathbf{y} = \pi(\mathbf{x})$.

When the projection is combined with a bistable mechanism (threshold, phase transition, precipitation), the output snaps to discrete values:

$$S = \sigma(\pi(\mathbf{x})) \quad (2)$$

where $\sigma : \mathbb{R}^m \rightarrow \{s_1, s_2, \dots, s_k\}$ is the discretizing function.

3.2 Physical Mechanisms of Discretization

Boundaries implement discretization through well-understood physical mechanisms:

- **Membranes:** Permeation has sharp concentration thresholds; the output is which molecules crossed, not their exact concentrations
- **Mineral surfaces:** Adsorption is selective with threshold behavior; pattern formation is inherently discrete
- **Phase interfaces:** Partitioning between phases shows cooperative transitions; the oil/water distribution reflects bulk state through discrete pattern
- **pH indicators:** Sharp color transitions at threshold pH values
- **Precipitation:** Above solubility product, precipitate forms; below, solution remains clear

In each case, continuous internal chemistry produces discrete outputs because the *boundary physics* involves bistability or threshold behavior.

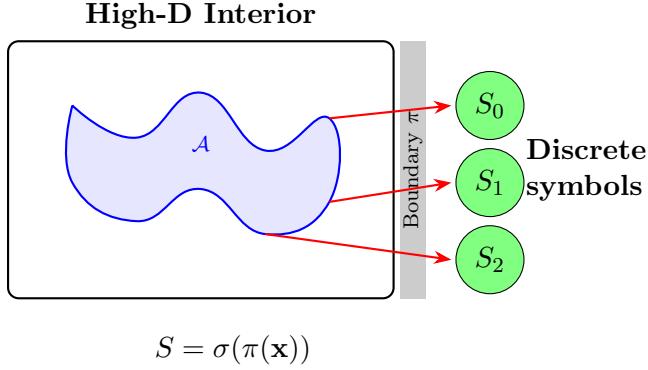


Figure 2: Continuous high-dimensional dynamics become discrete symbols at boundaries. The attractor manifold \mathcal{A} projects through boundary π onto distinguishable outputs.

3.3 Mathematical Basis

The emergence of discrete symbols from continuous dynamics via projection is a consequence of geometry. When a high-dimensional set projects onto a low-dimensional space, the push-forward density is generically non-uniform:

Observation (Projection induces structure). Let $\mathcal{A} \subset \mathbb{R}^n$ be a set traced by entrained dynamics (not necessarily a smooth manifold), and let $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear projection with $m < \dim(\mathcal{A})$. Then for typical π :

- (i) The projected density is non-uniform (clusters and gaps arise from curvature and self-overlap)
- (ii) Multiple interior states can map to the same boundary value (folds/aliasing)
- (iii) Sharp density gradients appear where attractor geometry changes orientation relative to the projection

These features—clusters, folds, boundaries—provide the raw material for discretization. A threshold mechanism σ converts the clustered projection into discrete symbols.

Capacity scaling (heuristic). The number of distinguishable symbols grows with the dimensionality mismatch. Numerical simulation of coupled oscillator networks suggests rapid growth as the ratio d/m increases, consistent with concentration-of-measure intuitions: high-dimensional distributions concentrate near shells, and projections sample these shells at many distinct angles. The key point is qualitative: high-dimensional attractors ($d \gg 1$) projected onto low-dimensional readouts ($m \sim 1$) produce many distinguishable symbols.

4 The Transition to Life

4.1 Stage 1: External Organization

Initially, prebiotic chemistry tracks an *external* attractor—the solar/tidal cycle. The proto-codes that emerge are:

- Determined by external forcing
- Reproducible under same forcing
- History-dependent (initial conditions select position on attractor)

The chemistry exhibits information-like behavior, but it is organized *by* the environment, not *by itself*.

4.2 Stage 2: Compartmentalization

Lipid bilayers form spontaneously under prebiotic conditions (Deamer and Oro, 1980). Encapsulation produces:

- **Concentration:** Interior chemistry is more concentrated than bulk

- **Defined boundary:** The membrane becomes a specific readout interface
 - **Partial decoupling:** Interior dynamics can develop independently of bulk
- The compartment creates conditions for a qualitative transition.

4.3 Stage 3: Endogenous Attractor

The key transition: inside the compartment, some chemistry becomes **autocatalytic**—self-amplifying reaction networks that can sustain themselves (Kauffman, 1986; Hordijk and Steel, 2004). Such networks, once established, can generate their own dynamics independent of external forcing (England, 2013).

Now the interior dynamics can generate their own attractor, independent of external forcing.

The transition to life: Swap the external attractor (sun-tracking) for an endogenously generated attractor (self-organization). The “code” is now self-maintained.

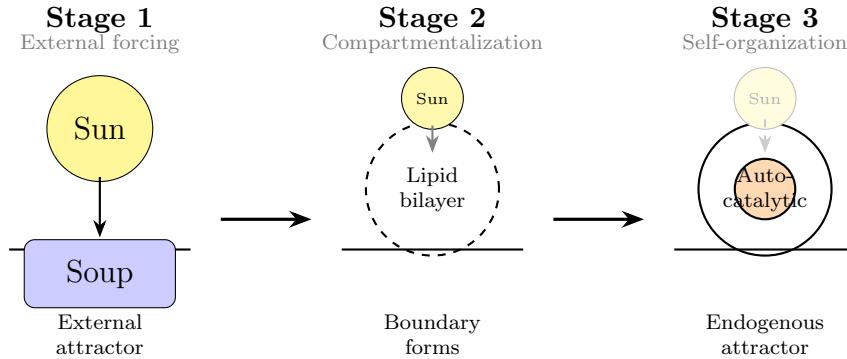


Figure 3: The transition to life. Stage 1: Chemistry tracks external attractor (sun). Stage 2: Compartmentalization creates defined boundary. Stage 3: Internal autocatalytic network generates its own attractor; external forcing becomes optional (faded sun).

This framework dissolves several puzzles:

- **Why codes?** Not designed—emerged from projection
- **Why discrete?** Boundary physics, not chemistry
- **Why self-maintaining?** Attractor swap from external to internal

5 Experimental Implications

5.1 Testing the Framework

The framework makes specific predictions testable in laboratory settings:

1. **Diversity requirement:** More chemical species → higher-D attractor → more distinguishable boundary states. This can be tested by varying mixture complexity.
2. **Forcing requirement:** Common forcing is necessary for attractor formation. Without external rhythm, the chemistry equilibrates and loses history-dependence.
3. **Boundary discretization:** Different boundary mechanisms (membrane, surface, phase interface) should produce different symbol alphabets from the same interior dynamics.
4. **Reproducibility:** Same initial conditions + same forcing → same boundary state. This is the operational definition of “code.”

5.2 Experimental Design

A minimal test system would include:

Substrate: 20–30 prebiotically plausible species (amino acids, sugars, nucleobases, metal ions, phosphate, fatty acids) in aqueous buffer.

Forcing: UV cycling (simulating day/night) + temperature cycling (20–60°C) over periods of hours to days.

Boundary: Lipid vesicles or mineral surface (montmorillonite clay).

Readout: pH indicator dyes or precipitation assays providing discrete outputs.

Protocol:

1. Define 32–64 input conditions (forcing phase, initial pH, metal ratios)
2. Run system for 10–100 forcing cycles
3. Record boundary state at end of each cycle
4. Cluster outputs to discover emergent symbols
5. Measure reproducibility (same input → same output)

The prediction is that ≥ 32 distinguishable boundary states emerge without being designed, with reproducibility $> 70\%$.

6 Discussion

6.1 Prior Experimental Work

Several experimental traditions have explored related phenomena without the specific framing of boundary-projected code emergence:

Oscillating chemical reactions. Belousov-Zhabotinsky (BZ) systems and coupled oscillator networks exhibit discrete phase states and synchronization Epstein and Showalter (1996). Showalter and colleagues demonstrated pattern formation and rudimentary computation in BZ droplet arrays. However, this work treats phase states as *computational* rather than as emergent *codes*—no encoding tables are populated, and the question “how many distinguishable symbols?” is not asked.

Protocell experiments. Szostak’s laboratory demonstrated that lipid vesicles can grow, divide, and selectively concentrate reactants Szostak et al. (2001). Deamer showed boundary-mediated chemistry in prebiotic conditions Deamer and Oro (1980). These experiments establish that compartmentalization affects chemical dynamics, but focus on *replication* and *metabolism* rather than *information encoding*. No input→output mappings are measured.

Autocatalytic networks. Kauffman’s theoretical work on autocatalytic sets Kauffman (1993), formalized by Hordijk and Steel Hordijk and Steel (2004), shows that self-sustaining reaction networks emerge spontaneously above a complexity threshold. Experimental validation exists for simple RAF (reflexively autocatalytic and food-generated) sets. This addresses “how does metabolism bootstrap?” but not “how do discrete codes emerge from continuous chemistry?”

Assembly theory. Cronin and colleagues developed assembly theory to quantify molecular complexity and demonstrated that complex molecules can form through iterated assembly Cronin et al. (2017). This addresses molecular *complexity* but not informational *discretization*.

Our distinction. We propose a specific *architecture*—high-D substrate, periodic forcing, boundary projection, bistable readout—and a specific *measurement*—encoding/decoding tables populated empirically. The claim is that discrete codes emerge from this architecture as a matter of physics, not that complex chemistry exists (which is well established).

6.2 Relation to Existing Theories

Our framework complements rather than contradicts existing origin-of-life theories:

- **RNA World:** RNA could be one of many molecular species tracking external forcing; its eventual dominance reflects its autocatalytic capabilities (Stage 3 transition), not a privileged role in code generation
- **Metabolism-first:** Autocatalytic metabolic cycles are exactly the “endogenous attractors” of Stage 3; our framework explains why discrete codes accompany them
- **Lipid World:** Compartmentalization is Stage 2—necessary for the transition but not sufficient; the attractor swap is the key step

6.3 Information Without Observers

A subtle point: in our framework, “information” exists without observers. The boundary states are discrete whether or not anyone reads them. This objective information—distinguishable, reproducible states—is what evolution later learns to use.

This resolves debates about whether information is “real” or merely observer-dependent. Emergent digitality is a physical phenomenon: high-D dynamics projected through bistable boundaries produce discrete states as a matter of physics, not interpretation.

6.4 Implications for Astrobiology

If codes emerge generically from diverse chemistry + forcing + boundaries, then:

- Life-like information processing may be common wherever these conditions obtain
- The specific chemistry matters less than the architecture (high-D, forced, bounded)
- Detection of extraterrestrial “life” might focus on boundary signatures rather than specific molecules

7 Conclusion

We have argued that the origin of biological codes is not a problem of engineering but of recognition. Discrete, symbolic information emerges inevitably when:

1. High-dimensional chemistry tracks common external forcing (creating attractor manifolds)
2. Boundaries project this high-D dynamics onto low-D observables (dimensional reduction)
3. Bistable physical mechanisms discretize the projected outputs (threshold physics)

The transition to life occurs when compartmentalization allows chemistry to swap external attractors for endogenous ones—self-organization replacing sun-organization.

This framework has several virtues:

- **Genericity:** No special chemistry required; any diverse mixture works
- **Inevitability:** Codes emerge from physics, not design
- **Testability:** Specific predictions about diversity, forcing, and boundary effects
- **Parsimony:** Explains discreteness without invoking “digital chemistry”

The genetic code was not invented. It was discovered—at the interface between complex chemistry and simple readouts, where high-dimensional continuous dynamics inevitably project into discrete symbols.

References

- Chyba, C. and Sagan, C. (1992). Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature*, 355(6356):125–132.
- Deamer, D.W. and Oro, J. (1980). Role of lipids in prebiotic structures. *BioSystems*, 12(3-4):167–175.

- Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523.
- England, J.L. (2013). Statistical physics of self-replication. *Journal of Chemical Physics*, 139(12):121923.
- Epstein, I.R. and Showalter, K. (1996). Nonlinear chemical dynamics: oscillations, patterns, and chaos. *Journal of Physical Chemistry*, 100(31):13132–13147.
- Hordijk, W. and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461.
- Kauffman, S.A. (1986). Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1–24.
- Kauffman, S.A. (1993). *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press.
- Miller, S.L. (1953). A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529.
- Pross, A. (2012). *What is Life? How Chemistry Becomes Biology*. Oxford University Press.
- Ruiz-Mirazo, K., Briones, C., and de la Escosura, A. (2014). Prebiotic systems chemistry: new perspectives for the origins of life. *Chemical Reviews*, 114(1):285–366.
- Schrödinger, E. (1944). *What is Life?* Cambridge University Press.
- Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001). Synthesizing life. *Nature*, 409(6818):387–390.
- Cronin, L., et al. (2017). Quantifying the origins of life on a planetary scale. *Proceedings of the National Academy of Sciences*, 114(49):13202–13207.
- Walker, S.I. and Davies, P.C.W. (2013). The algorithmic origins of life. *Journal of the Royal Society Interface*, 10(79):20120869.