The Code-Constraint Problem in Biological Systems: How Low-Dimensional Interfaces Shape High-Dimensional Dynamics

Ian Todd

Sydney Medical School

University of Sydney

Sydney, NSW, Australia

itod2305@uni.sydney.edu.au

December 20, 2025

Abstract

Biological systems operate in high-dimensional state spaces—protein conformational ensembles, gene regulatory networks, neural population dynamics—yet experimentalists routinely characterize them through low-dimensional observables: order parameters, principal components, expression markers. This dimensional mismatch creates a fundamental problem: when does a low-dimensional readout faithfully represent the underlying system, and when does it systematically distort our understanding? This perspective synthesizes insights from dynamical systems theory, statistical mechanics, and information theory to propose a unifying framework: low-dimensional codes function as *stabilizing constraints* rather than predictive representations. Using coupled oscillator simulations as a minimal model, we demonstrate that bandwidth-limited coupling induces systematic complexity collapse in responding systems while

maintaining bounded tracking—a signature distinct from information loss or prediction failure. The constraining effect requires structured projections that capture coherent collective variables; random projections of the same dimensionality fail. We connect this mechanism to concrete biophysical contexts: protein folding landscapes, single-cell state manifolds, morphogenetic fields, and neural coding. The framework suggests that biological coding structures—from genetic regulatory motifs to homeostatic setpoints—may function primarily as dimensional constraints that enable persistent organization, rather than as channels optimized for information transmission. This perspective offers practical criteria for distinguishing when coarse-grained models are reliable versus when they become "shadows" of latent dynamics.

Keywords: Dimensional reduction, coarse-graining, biological codes, constraint dynamics, order parameters, model identifiability

1 The Dimensional Mismatch Problem

Modern biology confronts a general tension. The systems we study—proteins, cells, tissues, neural circuits—operate in state spaces of enormous dimensionality. A modest protein explores a conformational landscape with thousands of degrees of freedom. A single cell's transcriptome spans tens of thousands of genes. A cortical column contains millions of synapses. Yet our experimental readouts are inevitably low-dimensional: we measure order parameters, principal components, marker genes, population firing rates.

This dimensional mismatch is not merely a practical limitation—it creates systematic interpretive challenges:

- **Model identifiability**: Parameter estimation in high-dimensional systems is generically "sloppy"—many parameter combinations produce indistinguishable low-dimensional outputs (Gutenkunst et al., 2007; Transtrum et al., 2015).

- **Projection artifacts**: Dimensionality reduction techniques (PCA, t-SNE, UMAP)

can create spurious clusters, collapse distinct states, and distort neighborhood relationships (Chari and Pachter, 2023; Kobak and Berens, 2019).

- **Observables vs. latent dynamics**: Order parameters that appear to govern dynamics may be shadows of higher-dimensional processes, leading to apparent "downward causation" that dissolves under finer measurement (Flack, 2017; Balduzzi and Tononi, 2008).

These problems span biophysics, systems biology, and neuroscience. When can we trust a low-dimensional model? When is dimensional reduction a faithful compression versus a systematic distortion?

1.1 A Unifying Perspective: Codes as Constraints

This paper proposes a framework for understanding dimensional mismatch in biological systems. The core insight is that low-dimensional interfaces between coupled systems function as *stabilizing constraints* rather than information channels.

Consider any biological interface: a cell membrane transducing environmental signals, a morphogenetic gradient coordinating tissue development, a neural code coupling sensory input to motor output. In each case, a high-dimensional "driving" system couples to a high-dimensional "responding" system through a low-dimensional bottleneck. The question is: what does this bottleneck *do*?

The standard information-theoretic framing asks how much information about the driving system's state is preserved or lost. We propose a different question: how does the bottleneck *shape the dynamics* of the responding system?

Using a minimal model of coupled oscillator lattices, we demonstrate that bandwidth-limited coupling produces a distinctive signature:

1. **Complexity collapse**: The responding system's effective dimensionality decreases systematically with code bandwidth.

2. **Bounded tracking**: Alignment between systems remains stable despite information loss.

3. **Structure dependence**: The effect requires projections onto coherent collective variables; random projections fail.

This pattern—complexity collapse with bounded tracking—is qualitatively different from information loss (which would produce tracking failure) or simple filtering (which would not systematically reduce effective dimensionality). We argue it represents a general mechanism by which biological coding structures enable persistent organization.

Positioning relative to existing theory: The framework relates to Haken's synergetics, where fast variables become "slaved" to slow order parameters—but extends this to coupled systems where the order parameter constrains a *separate* responding system. It differs from the information bottleneck principle, which optimizes compression for predictive relevance; we instead measure how compression constrains downstream dynamics, regardless of predictive content. It complements Markov blanket formalism by adding bandwidth as a quantitative knob: not just whether states are separated, but how many degrees of freedom can cross the boundary.

To be explicit about the novel claim: we are not proposing a better compression or a better predictor. We propose a *distinct dynamical regime diagnostic*: complexity collapse in the responding system while tracking error remains bounded, where "bounded" means alignment with the *coarse-grained reconstruction*, not full state reconstruction. This regime appears only for structured projections (coherent collective variables) rather than arbitrary low-dimensional reductions.

2   Biophysical Contexts

Before presenting the formal model, we ground the framework in concrete biological systems where dimensional mismatch is experimentally relevant.

2.1 Protein Conformational Ensembles

Proteins exist as dynamic ensembles exploring rugged energy landscapes (Frauenfelder et al., 1991; Henzler-Wildman and Kern, 2007). The full conformational state space has thousands of dimensions, yet function is often characterized by a handful of order parameters: radius of gyration, native contacts, collective mode amplitudes.

The question of when order-parameter descriptions are reliable has been extensively studied (Best and Hummer, 2005; Noé and Clementi, 2017). Reaction coordinates that appear to govern folding kinetics may hide multiple parallel pathways (Pande et al., 2010). The "folding funnel" picture (Dill and Chan, 1997) is itself a low-dimensional projection that may obscure high-dimensional heterogeneity.

Our framework suggests a reinterpretation: rather than asking whether order parameters *represent* the ensemble, we ask whether they *constrain* accessible trajectories. A well-chosen collective variable doesn't merely summarize the ensemble—it restricts what conformational dynamics remain possible. This is the difference between a thermometer (which measures temperature without affecting it) and a thermostat (which constrains the system to a temperature range).

Translating the model: In protein-ligand binding, System A is the ligand's conformational dynamics in solution (high-dimensional, exploring many states). System B is the binding-pocket dynamics of the target protein. The "code" is the subset of ligand conformations that can physically dock—perhaps characterized by 2–5 collective variables describing complementary surface geometry. When ligand binds, these few degrees of freedom constrain the pocket's accessible conformations, inducing complexity collapse in the binding site while maintaining the lock-and-key alignment. The signature would be: (1) reduced conformational entropy of the bound pocket relative to apo form, (2) stable binding geometry despite ligand fluctuations, and (3) this constraint failing for decoy ligands that match steric volume but not the structured collective variables of true binding. One could test this by comparing crystallographic B-factors (conformational flexibility) of binding pockets with cognate

versus non-cognate ligands: the framework predicts that cognate binding induces greater complexity collapse.

2.2 Single-Cell State Manifolds

Single-cell RNA sequencing reveals that cell states occupy low-dimensional manifolds embedded in high-dimensional gene expression space (Trapnell, 2015; Weinreb et al., 2018). Pseudotime analyses, trajectory inference, and RNA velocity all rely on this manifold structure.

Yet recent critiques highlight systematic artifacts: t-SNE and UMAP can create spurious clusters, apparent trajectories may reflect technical noise, and neighborhood relationships are often distorted (Chari and Pachter, 2023; Cooley et al., 2019). The fundamental issue is that $\sim$20,000-dimensional expression space is projected onto 2–3 visualization dimensions, a compression ratio that inevitably loses information.

The constraint perspective offers a different question: which expression programs *constrain* cell fate transitions? A gene regulatory motif that restricts accessible trajectories (like a developmental "valve") functions differently from one that merely correlates with cell state. The former shapes dynamics; the latter reflects them.

2.3 Morphogenetic Fields and Bioelectric Gradients

Developmental biology increasingly recognizes that tissues coordinate through field-like gradients—morphogen concentrations, bioelectric potentials, mechanical stresses—that are inherently low-dimensional compared to the cellular heterogeneity they organize (Levin, 2021; Turing, 1952).

Michael Levin's work on bioelectric control of morphogenesis demonstrates that manipulating membrane voltage patterns can reprogram tissue-scale outcomes (Levin, 2021, 2014). The voltage gradient is a low-dimensional interface between high-dimensional cellular dynamics and tissue-scale organization. It functions as a constraint: cells "listen" to the field

and restrict their behaviors accordingly.

This is precisely the mechanism we formalize below: a high-dimensional responding system (individual cell dynamics) couples to a driving system (the broader tissue/environment) through a low-dimensional code (bioelectric gradients), and the code's bandwidth constrains the responding system's behavioral complexity.

2.4 Neural Population Codes

Neuroscience faces the dimensional mismatch problem acutely. Neural populations fire in high-dimensional pattern spaces, yet behavior is low-dimensional and motor outputs are constrained by biomechanics. How does the brain bridge this gap?

Population coding theories typically frame neural activity as *representing* task variables (Georgopoulos et al., 1986; Pouget et al., 2000). An alternative framing asks which aspects of population activity *constrain* downstream processing. The brain's extensive dimensionality reduction—from millions of cortical neurons to a few hundred motor neurons—is not merely information compression but trajectory constraint.

Recent work on neural manifolds supports this view: population activity is confined to low-dimensional subspaces that constrain what computations are possible (Gallego et al., 2017; Sadtler et al., 2014). Learning a new behavior requires either operating within the existing manifold or expanding it—a constraint on behavioral complexity, not just a summary of neural activity.

3 A Minimal Model of Code-Constraint Dynamics

We now formalize the constraint mechanism using a minimal model: two coupled lattices of phase oscillators where interaction is restricted to a bandwidth-limited projection.

3.1 Model Architecture

Consider two systems, each a one-dimensional lattice of N locally coupled phase oscillators. System A (the "driving" system) evolves autonomously:

$$\dot{\theta}_i^A = \omega_i + K \sum_{j \in \mathcal{N}(i)} \sin(\theta_j^A - \theta_i^A) + \eta_i^A(t), \tag{1}$$

where ω_i are intrinsic frequencies, K is coupling strength, $\mathcal{N}(i)$ denotes nearest neighbors, and $\eta_i^A(t)$ is Gaussian noise.

System B (the "responding" system) couples to A only through a low-dimensional code:

$$\dot{\theta}_i^B = \omega_i + K \sum_{j \in \mathcal{N}(i)} \sin(\theta_j^B - \theta_i^B) + \lambda \sin(\hat{\theta}_i^A - \theta_i^B) + \eta_i^B(t), \tag{2}$$

where $\hat{\theta}^A$ is a bandwidth-limited reconstruction of A's state.

The code C_k consists of the first k Fourier modes of A's phase field:

$$C_m = \frac{1}{N} \sum_{i=1}^{N} e^{i\theta_i^A} e^{-i2\pi mi/N}, \quad m = 0, \ldots, k. \tag{3}$$

This architecture captures the essential structure: two high-dimensional systems coupled through a low-dimensional interface. The bandwidth k is a tunable parameter controlling code dimensionality.

3.2 Quantifying Complexity and Tracking

We measure *effective dimensionality* via spectral entropy of the Fourier amplitude spectrum. Specifically, the amplitude $|C_m|$ of each Fourier mode is computed, normalized to form a probability distribution $p_m = |C_m|/\sum_m |C_m|$, and the Shannon entropy $H = -\sum_m p_m \log p_m$ is calculated. The effective dimensionality is then $N_{\text{eff}} = e^H$, which gives the *effective support size*—the number of modes that contribute substantially to the dynamics, weighted by their

relative amplitudes. This is not a topological dimension but a complexity measure analogous to the participation ratio in localization theory. The same qualitative results obtain using alternative metrics (e.g., participation ratio of the power spectrum, or counting modes above a threshold).

Tracking quality is measured as mean circular distance between systems: $\Delta = N^{-1} \sum_i |\sin((\theta_i^A - \theta_i^B)/2)|$.

4 Results: The Constraint Signature

4.1 Complexity Collapse Under Bandwidth Reduction

Figure 1 shows the central result. As code bandwidth k decreases:

- System B's effective dimensionality decreases systematically ($N_{\text{eff}} = 16.7 \rightarrow 10.9$)

- System A's complexity remains constant ($N_{\text{eff}} \approx 17.0$)

- The relationship is approximately monotonic

The bottleneck constrains only the responding system; the driving system's dynamics are unaffected by how it is observed. This asymmetry is the hallmark of constraint rather than mutual information loss.

4.2 Bounded Tracking Despite Complexity Collapse

Mismatch between systems increases only modestly as bandwidth decreases (Figure 2): $\Delta = 0.38 \rightarrow 0.50$, a $\sim$32% change, compared to $\sim$35% reduction in $N_{\text{eff}}(B)$.

This asymmetry is crucial. If the code merely lost information, we would expect commensurate tracking failure. Instead, B tracks *the code* faithfully—it aligns with A's coarse-grained representation. The code constrains B's behavioral complexity without compromising alignment.
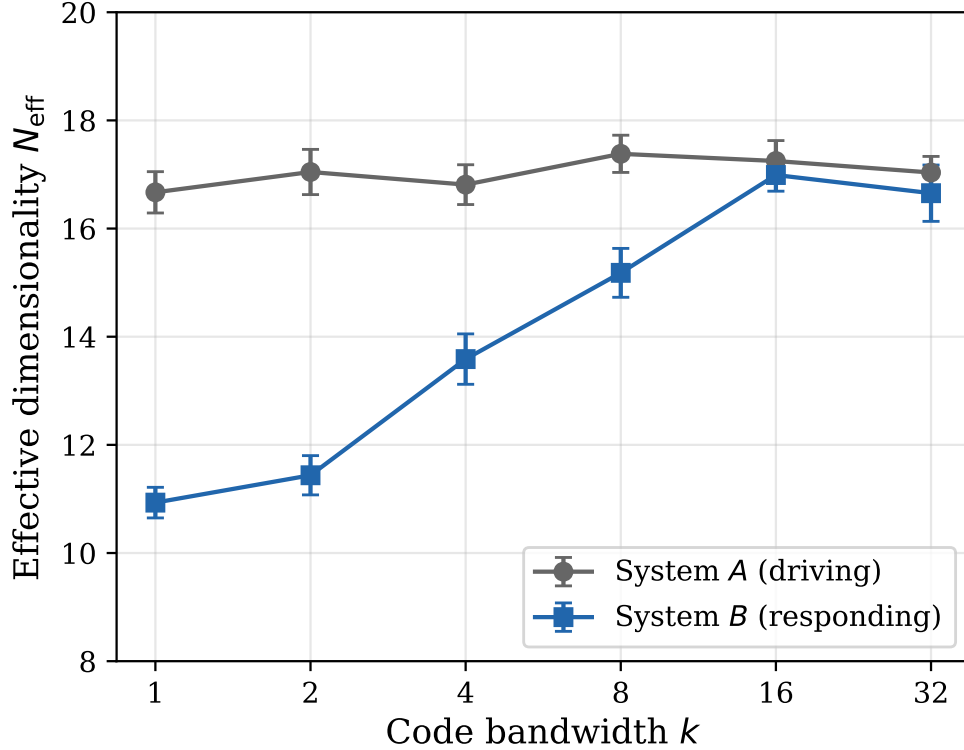
Figure 1: Effective dimensionality as a function of code bandwidth. System A (gray) maintains constant complexity; system B (blue) exhibits systematic collapse. Error bars: standard error over 15 trials.

4.3 Structure Matters: Random Projections Fail

A critical control: we repeated experiments using random k-mode projections instead of the lowest k Fourier modes. Results differ strikingly (Figure 3):

- Random projections produce *no* complexity collapse ($N_{\mathrm{eff}} \approx 15$–17)

- Mismatch is substantially higher

- At low k, random codes increase B's complexity above A's

The constraining effect is not about dimensionality per se, but about *structure*. Low-frequency modes capture spatially coherent patterns; random modes mix signal with noise. Biological codes, we argue, must similarly capture coherent collective variables rather than arbitrary dimensional reductions.
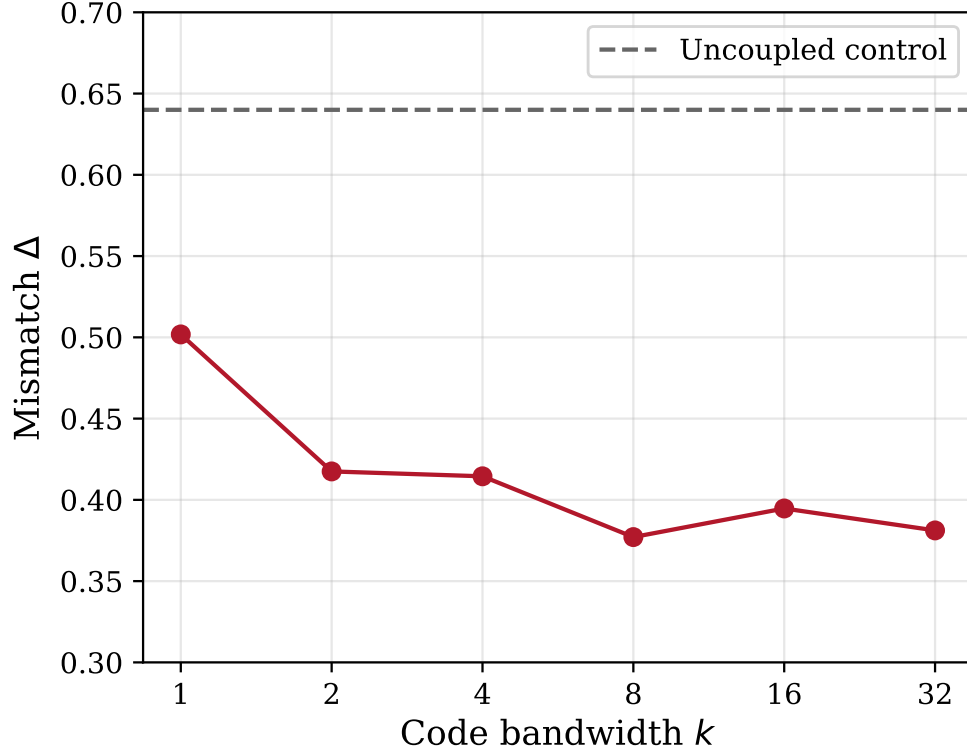
Figure 2: Mismatch between systems as a function of code bandwidth. Coupled systems (solid) maintain lower mismatch than uncoupled controls (dashed) across all bandwidths.

Notably, at low k the random projection actually *increases* B's complexity above baseline ($N_{\text{eff}} \approx 18$ vs. 17 for the uncoupled system). This occurs because sparse random projections destroy local spatial correlations in the driving signal, effectively acting as a whitening filter that forces the responding lattice to desynchronize locally while still responding to the global coupling. The random code injects high-frequency structure rather than removing it.

Robustness: The qualitative pattern—complexity collapse for structured projections, not for random ones—persists across the parameter ranges tested: noise amplitude $\sigma \in [0.1, 0.5]$, coupling strength $\lambda \in [0.25, 4.0]$, and lattice sizes $N \in [32, 128]$. The effect is not an artifact of specific parameter choices. We expect other structured bases (Laplacian eigenmodes, wavelets, principal components of the driving system) to produce similar constraint signatures, though the specific relationship between k and N_{eff} will depend on the basis.
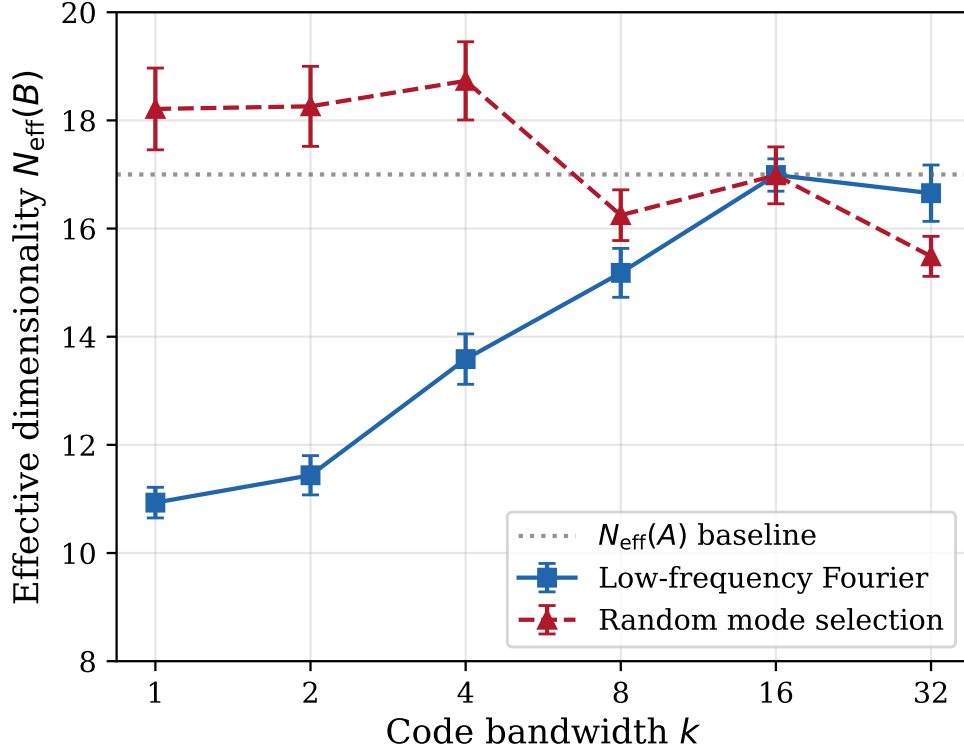
Figure 3: Low-frequency Fourier coupling (blue) vs. random mode selection (red). Random projections of the same dimensionality fail to induce complexity collapse.

5 Implications for Biological Coding Structures

5.1 Codes as Dimensional Valves

The simulation results suggest a general principle: biological interfaces that couple high-dimensional systems through low-dimensional projections function as *dimensional valves—* they constrain the behavioral complexity of responding systems without requiring full state reconstruction.

This reframes the function of biological coding structures:

- **Gene regulatory motifs** don't merely "represent" cellular state; they constrain accessible developmental trajectories.

- **Morphogenetic gradients** don't merely "signal" position; they restrict the dynamical complexity of responding tissues.

- **Neural codes** don't merely "encode" stimuli; they constrain downstream processing to behaviorally relevant dimensions.

The key insight is that constraint and representation are distinct functions. A code can effectively constrain dynamics while having poor predictive power—indeed, discarding destabilizing microstate detail may be precisely why the code works.

5.2   When Coarse-Grained Models Are Reliable

The framework offers practical criteria for evaluating low-dimensional models:

1. **Does the projection capture coherent collective variables?** Random or arbitrary projections will not produce stable constraint. Effective codes must select dynamically relevant degrees of freedom.

2. **Is there a responding system being constrained?** If the "code" is merely an observer's summary with no causal role, it may misrepresent the underlying dynamics. Constraint requires causal coupling.

3. **Does complexity collapse predict stability?** If reducing the code's bandwidth produces chaotic rather than simplified dynamics in the responding system, the projection is not capturing constraining structure.

These criteria distinguish cases where order-parameter descriptions are reliable (because they capture genuine constraints) from cases where they are misleading (because they are arbitrary projections onto a richer dynamics).

5.3   The Membrane as Paradigm Case

The cell membrane exemplifies this framework. It is a physical dimensional bottleneck: the external environment and internal cytoplasm are each enormously high-dimensional,

yet coupling occurs through a restricted set of interface variables—receptor occupancy, ion channel conductance, transmembrane voltage.

The membrane does not "represent" the environment; it *constrains* which environmental features can influence cellular dynamics. Evolution has tuned both the bandwidth (interface dimensionality) and structure (which degrees of freedom couple) to maximize actionable constraint per unit metabolic cost.

This perspective suggests that "information processing" metaphors may be misleading for membrane biology. The membrane is not a channel that transmits environmental information; it is a valve that restricts cellular dynamics to viable trajectories given environmental context.

6 Relation to Existing Theory

Our framework connects to several established theoretical traditions:

Synergetics and slaving: Haken's synergetics describes how fast variables become "slaved" to slow order parameters (Haken, 1983). We extend this to coupled systems: the code is a bandwidth-limited order parameter that constrains a separate responding system.

Information bottleneck: The information bottleneck principle (Tishby et al., 1999) optimizes compression for predictive relevance. Our framework differs: we measure how compression constrains downstream dynamics, not how much predictive information is preserved.

Markov blankets: The Markov blanket formalism (Friston, 2013) identifies boundaries separating internal from external states. We add a quantitative dimension: bandwidth k specifies how many degrees of freedom can cross the blanket.

Computational mechanics: ε-machines characterize minimal sufficient statistics for prediction (Crutchfield and Young, 1989). We propose a sibling concept: *minimal stabilizing interfaces*—projections that constrain dynamics while potentially being poor predictors.

7 Limitations and Future Directions

We emphasize what the simulation demonstrates directly: complexity collapse and bounded tracking under bandwidth-limited coupling. We hypothesize, but do not directly test, that dimensionality reduction is a prerequisite for attractor formation and persistent organization. Perturbation studies would be needed to verify resilience of the constrained dynamics.

The Fourier basis is convenient analytically but not uniquely privileged. The key property is capturing spatially coherent structure; other structured bases (wavelets, Laplacian eigenfunctions) might serve equally well. The critical distinction is between coherent projections that enable constraint and incoherent projections that do not.

Future work should test these predictions in experimental systems: does manipulating interface bandwidth (e.g., via optogenetic control of receptor expression) produce the predicted complexity collapse in responding systems?

8 Supplementary Results: Generality Across Scales and Dynamics

To test the generality of the constraint signature, we extended the simulations in two directions: larger system sizes and alternative dynamics.

8.1 Large-Scale Kuramoto Systems

We repeated the main analysis at $N = 512$ and $N = 1024$ oscillators (Supplementary Figure S1). The complexity collapse signature scales appropriately:

- At $N = 512$: $N_{\text{eff}}(B)$ ranges from ≈ 102 at $k = 4$ to ≈ 143 at $k = 128$

- At $N = 1024$: the complexity ratio $N_{\text{eff}}(B)/N_{\text{eff}}(A)$ ranges from 0.64 to 0.95 as k increases

- Random projections continue to show no collapse, with $N_{\text{eff}}(B)$ remaining near the uncoupled baseline

The constraint signature is not a finite-size artifact; it persists as system dimensionality increases.

8.2 Gene Regulatory Network Dynamics

To test whether the effect depends on oscillatory dynamics, we implemented a gene regulatory network (GRN) model with Hill-function activation:

$$\dot{x}_i = -\frac{x_i}{\tau} + \sigma\left(\sum_j W_{ij}x_j\right) + \eta_i(t), \tag{4}$$

where $\sigma(\cdot)$ is a sigmoid activation function, W is a sparse random connectivity matrix, and τ is the decay timescale.

The "code" in this context represents transcription factor readout: a low-dimensional projection (via DCT) of one network's expression state constrains the other network's dynamics.

Results (Supplementary Figure S2, $N = 256$ genes):

- $N_{\text{eff}}(A)$ remains constant at ≈ 193

- $N_{\text{eff}}(B)$ collapses from 192 at $k = 1$ to 121 at $k = 32$ (37% reduction)

- Mismatch increases modestly (0.22 to 0.37) as bandwidth decreases

The GRN results demonstrate that complexity collapse under bandwidth-limited coupling is not specific to phase oscillators. The same signature emerges in a fundamentally different dynamical system—one that more directly models gene regulatory interactions. This supports the claim that the constraint mechanism is general, not an artifact of Kuramoto dynamics.

9   Conclusion

We have proposed a framework for understanding dimensional mismatch in biological systems. Low-dimensional interfaces between coupled high-dimensional systems function as stabilizing constraints rather than information channels. The signature of effective constraint is complexity collapse with bounded tracking: the responding system's effective dimensionality decreases systematically with code bandwidth while maintaining alignment with the driving system's coarse-grained representation.

This framework offers a unifying lens for diverse phenomena—protein folding landscapes, single-cell manifolds, morphogenetic gradients, neural population codes—and provides practical criteria for evaluating when low-dimensional models are reliable versus when they become shadows of richer dynamics.

The key insight is that biological codes may function primarily as dimensional valves that enable persistent organization, rather than as channels optimized for information transmission. Constraint and representation are distinct functions, and conflating them may systematically mislead our understanding of biological organization.

Acknowledgments

Declaration of generative AI use in the writing process

During the preparation of this work the author used Claude (Anthropic) for manuscript drafting, editing, and code development. The author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

CRediT author statement

Ian Todd: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Data availability

All simulation code and data are publicly available at `https://github.com/todd866/code-formation-jt`

References

D. Balduzzi and G. Tononi. Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Computational Biology*, 4(6):e1000091, 2008.

R. B. Best and G. Hummer. Reaction coordinates and rates from transition paths. *Proceedings of the National Academy of Sciences*, 102(19):6732–6737, 2005.

T. Chari and L. Pachter. The specious art of single-cell genomics. *PLoS Computational Biology*, 19(8):e1011288, 2023.

S. M. Cooley, T. Hamilton, E. J. Deeds, and J. C. J. Ray. A novel metric reveals previously unrecognized distortion in dimensionality reduction of scRNA-seq data. *bioRxiv*, page 689851, 2019.

J. P. Crutchfield and K. Young. Inferring statistical complexity. *Physical Review Letters*, 63 (2):105–108, 1989.

K. A. Dill and H. S. Chan. From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1):10–19, 1997.

J. C. Flack. Coarse-graining as a downward causation mechanism. *Philosophical Transactions of the Royal Society A*, 375(2109):20160338, 2017.

H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

K. Friston. Life as we know it. *Journal of the Royal Society Interface*, 10(86):20130475, 2013.

J. A. Gallego, M. G. Perich, L. E. Miller, and S. A. Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.

A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986.

R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10):e189, 2007.

H. Haken. *Synergetics: An Introduction*. Springer, 1983.

K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450(7172): 964–972, 2007.

D. Kobak and P. Berens. The art of using t-SNE for single-cell transcriptomics. *Nature Communications*, 10:5416, 2019.

M. Levin. Molecular bioelectricity: how endogenous voltage potentials control cell behavior and instruct pattern regulation in vivo. *Molecular Biology of the Cell*, 25(24):3835–3850, 2014.

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

F. Noé and C. Clementi. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Current Opinion in Structural Biology*, 43:141–147, 2017.

V. S. Pande, K. Beauchamp, and G. R. Bowman. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods*, 52(1):99–105, 2010.

A. Pouget, P. Dayan, and R. Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132, 2000.

P. T. Sadtler, K. M. Quick, M. D. Golub, S. M. Chase, S. I. Ryu, E. C. Tyler-Kabara, B. M. Yu, and A. P. Batista. Neural constraints on learning. *Nature*, 512(7515):423–426, 2014.

N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pages 368–377, 1999.

M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, 143(1):010901, 2015.

C. Trapnell. Defining cell types and states with single-cell genomics. *Genome Research*, 25 (10):1491–1498, 2015.

A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B*, 237(641):37–72, 1952.

C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, and A. M. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.