Goldbach-Lemoine Descent Graphs

Peter Upham
Independent researcher
`founder@threebons.ai`
ORCID: 0009-0005-7512-095X

Abstract

We introduce the *Goldbach–Lemoine Descent Graph*, a discrete dynamical object formed through iterated additive decomposition. Using witness pairs selected at each step from a constrained search window according to a fixed policy, we construct a directed acyclic graph rooted in the source integer. Unconstrained Descent Graphs exist wherever the Goldbach and Lemoine conjectures hold; we empirically verify the existence of our constrained variants for all integers up to 10^9. Over a set of randomly sampled integers between 5×10^4 and 5×10^8, and across five policy frameworks (four deterministic and one random null), we instantiate the graphs and probe several key properties. For all integers $N \geq 98$ in our tested range, the graphs—across policies—terminate at a common set of ten non-consecutive primes. Aggregating the graph-wide branching weights at each sink yields a composition vector that constitutes an *arithmetic signature* of N. After ILR embedding, these signatures form policy-dependent manifolds. Despite the abundance of additive representations, deterministic selection produces sharply differentiated low-dimensional geometry, modular concentration phenomena, and long-range serial dependence. Under randomized witness selection these structures disappear. The results demonstrate that in the context of additive partitions, arithmetic selection rules alone can induce coherent geometric and dynamical organization in integer-indexed data.

Mathematics Subject Classification (2020): 11P32 11A25 37B20 55N31

Contents

1 Introduction

Goldbach and Lemoine representations provide abundant additive decompositions. Rather than attempting proofs, we treat these representations as a generative resource: imposing a deterministic selection rule on admissible witnesses induces a descent operator on integers. Iterating this operator yields a finite directed acyclic graph (DAG) whose terminal prime weights encode a compositional signature.

The central question is structural rather than existential: how does deterministic witness selection interact with the prime landscape to shape the geometry of these signatures? We show that simple policies generate sharply differentiated geometric, modular, topological, and dynamical behavior, while randomized selection eliminates these effects.

Earlier work by the author proved that for any odd $N \geq 9$, the interval $[1, N]$ is covered by Y-smooth numbers and their complements with respect to N, provided $Y \geq \lceil N/3 \rceil$. The cover was shown to fail in the near-critical regime $N/5 < Y < N/3$ if and only if a constrained Diophantine equation of the form $p + 2q = N$ admits a solution with $p, q > N/5$ and both prime. This connection to Lemoine's conjecture motivated the present exploration: can iterated Lemoine–Goldbach decomposition, with witness selection restricted to the critical window, generate objects with rich and reproducible structure?

1.1 Summary of results

The answer turns out to be yes, and surprisingly so. Our main experimental findings are as follows.

Common sinks. For every integer $N \geq 98$ in our experiments (up to 5×10^8), and across all five policies (labeled QUARTER, CENTER, DOWN, UP, RANDOM), every Descent Graph terminates at the same ten primes: $\{2, 3, 5, 19, 29, 37, 47, 59, 73, 97\}$. This finite-sink exhaustion is a nontrivial rigidity phenomenon: many structurally different trees collapse to a small, universal terminal basis.

Arithmetic signatures. The row-normalized vector of sink masses constitutes a composition on the 9-simplex—an arithmetic fingerprint of the integer N that is both stable enough to exhibit geometry and sensitive enough to support classification.

Emergent geometry. After ILR transformation, the embedded signatures form structured manifolds whose geometry depends strongly on policy. One policy produces a nearly one-dimensional spine with low-dimensional transversal structure; another produces an archipelago of near-discrete islands; a third produces a higher-dimensional, more globular shape; a fourth generates compact and distinct low-moderate dimensional structure, and RANDOM yields a high-dimensional cloud.

Intrinsic dimension. Two independent estimators (kNN-MLE and correlation dimension) consistently yield policy-specific intrinsic dimensions ranging from as low as 1.9–2.7 (QUARTER) to as high as 7.4 for the RANDOM null (near the ambient dimension of 9), confirming that deterministic policies impose genuine geometric constraints. Within each policy, dimension gradually *decreases* with increasing N, consistent with deeper descent trees accumulating additional modular constraints (Section 7.4). The policy dimensionality ordering QUARTER $<$ CENTER $<$ DOWN $<$ UP $<$ RANDOM is preserved across four orders of magnitude (Figure 8).

Persistent homology. Computing persistent homology through H_2 reveals a dimensional hierarchy in the topology: QUARTER has sparse, long-lived H_1 cycles and essentially no H_2; CENTER supports abundant H_1 and robust H_2 (loops and voids); UP produces many short-lived features across all degrees. The topological complexity tracks intrinsic dimension in a manner consistent with theory.

Terminal basis refinement. The 10-sink system admits principled refinement of the terminal basis using targeted "early stopping" of the recursion. For exposition, a 12-sink example is instantiated. It preserves the qualitative geometry, confirming that the phenomena are not artifacts of the particular sink set.

Modular arithmetic skeleton. The raw integer-valued sink coefficients $W_s(N)$ carry modular structure. Three sinks—$\{2, 5, 59\}$—have *deterministically even* coefficients across all policies and all residue classes, a consequence of their unreachability along Lemoine p-chains (Proposition 1). Beyond parity, the QUARTER policy induces extreme mod-3 concentration at the large sinks ($\Pr[3 \mid W_{59}] \approx 0.77$), driven by a cancellation cascade in the Lemoine recurrence $W_s(M) \equiv W_s(p) - W_s(q) \pmod 3$. This modular collapse provides a mechanistic explanation for QUARTER's anomalously low intrinsic dimension (Section 8).

Short memory. Terminal sink mass is highly predictable from witness statistics at the first one or two decomposition steps. This short-memory property is strongest for DOWN and weakest for QUARTER, and invites comparison with Markov-like behavior in the induced state representation on the simplex.

Long memory. In the orthogonal "horizontal" direction—across consecutive integers—the sink-weight sequence under deterministic policies exhibits strong serial dependence. Shannon entropy stratifies sharply by residue class ($N \bmod 6$), with the governing modulus depending on policy: parity for DOWN, $N \bmod 3$ for QUARTER. A three-stage decomposition reveals three distinct layers of serial dependence: (i) strong mod-6 periodicity (deterministic, fully explained by residue-class structure); (ii) secular drift from the slow evolution of prime density, which inflates naïve Hurst estimates by $\approx$0.1–0.15; and (iii) residual persistence, with fully-detrended Hurst exponents $H \approx$ 0.76–0.86 across all four deterministic policies, confirmed at $Z > 18$ above a within-class shuffle null. A decisive *random-witness null test*—replacing deterministic witness selection with uniform random choice among valid witnesses—yields $H = 0.49$ (pure white noise), demonstrating that the persistence is created by the deterministic selection rule, not by the prime distribution. The scaling is monofractal ($\Delta\alpha < 0.06$), consistent with a single mechanism. The white-noise output of the random policy also raises an interesting equidistribution question about Lemoine witnesses (Question 1).

1.2 Related literature

The Goldbach conjecture has been verified computationally to 4×10^{18} [5]. Lemoine's conjecture (also known as Levy's conjecture [9]), asserting that every odd integer $M \geq 7$ can be written as $M = p + 2q$ with p, q prime, has been verified to 10^{13} [10]. The constrained window $q \in [\lceil M/5 \rceil, \lfloor M/3 \rfloor]$ used in this paper is motivated by the deterministic smooth-number covering theorem of [7], which proves that the interval $[1, N]$ is covered by $\lceil N/3 \rceil$-smooth numbers and their complements, and identifies the Lemoine equation $p + 2q = N$ with N odd (or $p + 2q = N + 1$ with N even), and $p, q > \lfloor N/5 \rfloor$ and prime as the unique obstruction in the near-critical regime.

Iterated number-theoretic maps have a long history as generators of discrete dynamical systems. The Collatz conjecture ($3n + 1$ problem) produces unary trajectories that are conjectured to terminate universally at the cycle $\{1, 2, 4\}$; aliquot sequences exhibit similarly complex dynamics with conjecturally universal bounded behavior. The Descent Graphs studied here differ from these classical constructs in that they are *branching*: each node may decompose into multiple children, producing a full directed acyclic graph rather than a linear trajectory.

Nomenclature. We note that the term "Goldbach graph" appears in the literature with a different meaning. Das, Ghosh, Ghosh, and Sen [13] study undirected bipartite "odd-even" graphs whose vertices are even integers and odd integers and whose edges encode prime-sum adjacency; their focus is on connectedness (equivalent to the Goldbach conjecture) and Hamiltonian properties. Separately, Fujiyoshi [14] introduces "additive prime trees" associated with finite subsets of primes and studies their finiteness, applying this to decompose spaces of cusp forms into products of Eisenstein series. Our Descent Graphs are fundamentally distinct from both: they are directed acyclic graphs generated by an *algorithmic recursion* governed by constrained Lemoine/Goldbach witness selection, not static adjacency structures or modular-form decomposition tools.

The geometric analysis relies heavily on the framework of Compositional Data Analysis (CoDA), developed by Aitchison [1], which provides the Isometric Log-Ratio (ILR) transformation used to map simplex-valued data into unconstrained Euclidean coordinates. CoDA has found extensive application in geology and genomics, but to our knowledge has not previously been applied to number-theoretic data.

For intrinsic dimension estimation, we employ the kNN-MLE method of Levina and Bickel [4] and the correlation dimension of Grassberger and Procaccia [3], both standard in the analysis of fractal and low-dimensional datasets. The persistent homology computations follow the Vietoris–Rips filtration framework of Edelsbrunner and Harer [2], implemented via Ripser [6]; the broader program of applying algebraic topology to data analysis is surveyed by Carlsson [11].

The long-memory analysis uses Detrended Fluctuation Analysis (DFA) [16], cross-validated by Rescaled Range (R/S) analysis [12].

Fractal statistics of Goldbach partitions. The fractal structure of the Goldbach partition function $r(n)$ (the number of representations of n as a sum of two primes) was investigated by Wang, Huang, and Dai [20], who documented prominent 3-period oscillations and a self-similar hierarchical decomposition of the $r(n)$ series. Separately, Shanker [21] applied Rescaled Range analysis to the spacings of Riemann zeta zeros, obtaining a Hurst exponent of $H \approx 0.095$ (strongly anti-persistent) and a fractal dimension of 1.9, stable across fifteen orders of magnitude. Our findings connect to both: the mod-3 and mod-6 entropy stratification documented in Sections 11.1 and 8.3 provides a plausible *mechanistic* explanation for the period-3 oscillations observed by Wang et al., while our DFA analysis extends the Hurst-exponent methodology to a compositional setting and isolates the deterministic selection rule—rather than the witness distribution itself—as the source of the observed persistence.

1.3 Outline

Section 2 gives precise definitions. Section 3 describes the five policy frameworks (four deterministic and one random null) and discusses existence. Section 4 presents the experimental design. Section 5 establishes the common sink set and introduces the 12-sink alternative. Section 6 develops the simplex geometry and ILR embedding. Section 7 estimates intrinsic dimension and documents its contraction across orders of magnitude. Section 8 analyzes the modular arithmetic structure of the raw sink coefficients, proving a parity skeleton theorem and identifying the mod-3 cancellation cascade that connects witness selection to manifold geometry. Section 9 presents the persistent homology analysis. Section 10 develops the algebraic framework underlying terminal basis refinement, proving basis equivalence and establishing the poset structure of admissible bases. Section 11 analyzes the memory structure of the descent dynamics, including vertical short memory (predictability from early decomposition steps), entropy stratification by residue class, spectral structure, and horizontal serial dependence via a three-stage Hurst exponent decomposition with permutation testing and a random-witness null test that identifies the deterministic selection rule as the source of persistence. Section 12 gathers open questions and conjectures.

2 Definitions and setup

2.1 Lemoine and Goldbach decomposition

Definition 1 (Lemoine search window)**.** For an odd integer $N \geq 7$, define the *constrained Lemoine search window* as the set of primes q satisfying

$$\lceil N/5 \rceil \;\leq\; q \;\leq\; \lfloor N/3 \rfloor$$

such that $p = N - 2q$ is also prime. A *Lemoine witness* for N in this window is any such pair (p, q).

The window bounds arise from the critical regime identified in the author's earlier work: at $N/3$ and above, a specific smooth number cover succeeds trivially. At $N/5$ and below, cover failures correspond to multiple Diophantine obstructions. The interesting arithmetic lives in between, where a failure of the cover is equivalent to only one obstruction: a Lemoine representation.

For an even integer $N \geq 4$, a *Goldbach witness* is a pair of primes (a, b) with $a + b = N$. By convention we scan from $a = N/2$ downward.

In both cases, the witness selection is not unique in general. The choice of which witness to use at each step is governed by a *policy*, defined next.

2.2 The Descent Graph

Definition 2 (Descent Graph)**.** Fix a policy π (defined in Section 3). Given an integer $N > 97$, the *Goldbach–Lemoine Descent Graph* $G_\pi(N)$ is the rooted DAG constructed as follows. (We use "DAG" in the sense that all edges are directed and the graph is acyclic; however, we do not merge nodes with the same integer value, so the structure is a rooted tree.) Place N at the root. If N is odd and $N > 97$, select the π-designated Lemoine witness (p, q) and add directed edges $N \to p$ and $N \to 2q$. If N is even, select the π-designated Goldbach witness (a, b) and add edges $N \to a$ and $N \to b$. Since Goldbach decomposes an even integer into two odd summands, each successive step decomposes odd integers into Lemoine witnesses. Recurse on each child that is not a terminal (sink). A node is a *sink* if it belongs to a fixed terminal set S (defined in Section 5). The recursion is guaranteed to terminate because each child is strictly smaller than its parent.

Each non-sink internal node has exactly two children (one from each summand). The *sink weight* $W_s(N)$ is defined recursively: for a Lemoine node $M = p+2q$, set $W_s(M) = W_s(p)+2\,W_s(q)$; for a Goldbach node $N = a + b$, set $W_s(N) = W_s(a) + W_s(b)$; and for a sink s', set $W_{s'}(s') = 1$ and $W_s(s') = 0$ for $s \neq s'$. Equivalently, $W_s(N)$ counts root-to-s paths weighted by the coefficient (1 or 2) inherited at each branching. The vector $\mathbf{w}(N) = (W_s)_{s \in S}$, normalized to sum to 1, is the *arithmetic signature* of N under policy π.

2.3 The dynamical systems viewpoint

Each policy π induces a map $F_\pi : \mathbb{N} \to \mathbb{N}^2$ sending each non-sink integer to its two children. By iteration, F_π generates the rooted DAG. The sink-coefficient vector $\mathbf{w}(N)$ is a projection of the full tree onto the fixed terminal basis S. After row normalization, $\mathbf{w}(N)$ lives on the $(|S| - 1)$-simplex Δ^9. Applying the ILR transform maps Δ^9 isometrically into $\mathbb{R}^9$, and the resulting point cloud $\{\mathrm{ILR}(\mathbf{w}(N)) : N \in \text{sample}\}$ becomes the object of geometric and dynamical analysis.

This framing—policy as control parameter, simplex as state space, ILR coordinates as Euclidean representation—unifies the rest of the paper. Different policies produce different dynamics on the same state space.

3 Policy frameworks

We study five policies, four deterministic and one random. Each deterministic policy defines a rule for selecting a witness within the Lemoine window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ (for odd N). Across all policies, the Goldbach scan (for even N) searches outward from N/2 for the first witness (odd+odd). Thenceforth, every subsequent decomposition generates odd-odd pairs, and thus the policy reverts to the odd N decomposition until termination.The random policy, which searches at each decomposition step from a random location in the near-critical window, serves as a null test.

3.1 down (N/3-down)

Scan q downward from $\lfloor N/3 \rfloor$, testing odd values, until both q and $p = N - 2q$ are prime. This selects the witness closest to the upper boundary of the window, i.e., the largest admissible q. DOWN tends to produce balanced decompositions (p and q of comparable size).

3.2 up (N/5-up)

Scan q upward from $\lceil N/5 \rceil$ until a valid witness is found. This selects the smallest admissible q, near the lower window boundary. UP tends to produce highly unbalanced decompositions (one large, one small child) and generates wider, bushier descent graphs.

3.3 quarter (N/4-outward)

Begin scanning from $q = \lfloor N/4 \rfloor$, the midpoint of the window, and alternate outward (checking $q-1, q+1, q-2, q+2, \ldots$) until a witness is found. QUARTER targets the geometric center of the window. Its embedding space has the lowest intrinsic dimension among all policies.

3.4 center (4/15-centered)

Begin at $q = 4N/15$, which lies equidistant between 1/5 and 1/3, and scan outward. The fraction 4/15 is one natural choice; the code supports arbitrary center fractions strictly between 1/5 and 1/3. CENTER interpolates between DOWN and QUARTER in its geometric properties.

3.5 random (uniform null)

For each odd node M in the descent tree, choose uniformly at random among all valid Lemoine witnesses in the constrained window. A per-root deterministic seed ensures reproducibility. The Goldbach decomposition at even nodes uses the same deterministic $\lfloor N/2 \rfloor$-down scan as all other policies, so RANDOM randomizes *only* the Lemoine witness selection—isolating a single degree of freedom.

The random policy serves a dual role. Structure that vanishes under random selection (e.g., long-range serial dependence, dimensional contraction) is attributable to the deterministic selection rule. Structure that *survives* (e.g., the parity skeleton, the reconstruction identity) reflects properties of the witness pool itself.

Remark 1 (The policy design space). The five policies above are all *search-start-location* rules: they differ only in where within the window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ the scan begins, then proceed by nearest-neighbor search. Because Lemoine witnesses are empirically abundant at the scales we study, the first valid witness is typically close to the starting point, so downstream structure is driven principally by the choice of starting location. However, this is a narrow slice of the space of admissible policies; the broader landscape is discussed in Question 2.

3.6 Existence

For unconstrained decompositions (the full Lemoine or Goldbach window), existence follows from the respective conjectures, both verified computationally to well beyond our range. Our constrained window $q \in [\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ is precisely the regime where only the $\{1, 2\}$-partition (the Lemoine equation) can produce an obstruction.

3.7 Heuristic abundance

A Hardy–Littlewood heuristic predicts that the number of constrained Lemoine witnesses in the window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ grows on the order of

$$\frac{N}{(\log N)^2}.$$

For example, at $N = 10^6$ the heuristic predicts roughly 4×10^2 witnesses, consistent with observed counts.

We verified computationally that for every integer $N \leq 10^9$ and every policy studied, at least one admissible witness exists in the constrained window. No failures were encountered.

4 Experimental design

We sampled integers from the range $[5 \times 10^4, 5 \times 10^8]$ using both uniform sampling (for global analysis of geometry) and stratified random sampling (for scale-based analysis). For our serial dependence experiments, we studied 100,000 consecutive integers ranging from $5,000,001 - 5,100,000$. For each sampled integer N and each of the four policies, we constructed the full Descent Graph, recording:

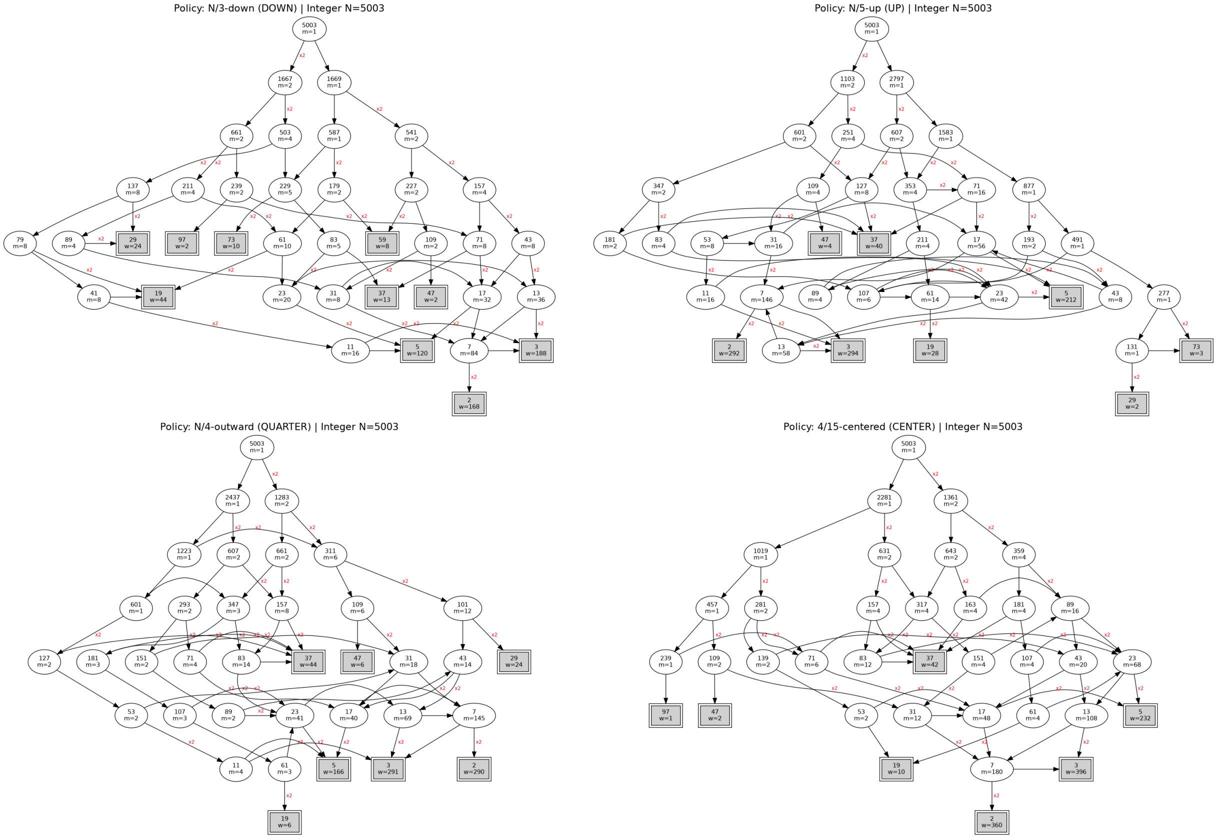
(a) the sink-weight composition vector $\mathbf{w}(N)$ (10 or 12 components, one per terminal prime);

(b) depth-1 and depth-2 witness traces: the specific (p, q) or (a, b) pairs chosen at the first two levels of recursion;

(c) depth-wise branching statistics: node counts, internal vs. sink counts, and cumulative mass flow at each depth level;

(d) graph-level summaries: maximum depth, total node count, and total internal node count.

All computations used deterministic primality testing (Miller–Rabin with bases $\{2, 3, 5, 7, 11\}$, provably correct for all $n < 2.15 \times 10^{12}$ and therefore certainly correct throughout our computational range) and were implemented in Python with memoized witness caching for efficiency.


Policy: N/3-down (DOWN) | Integer N=5003

Policy: N/5-up (UP) | Integer N=5003

Policy: N/4-outward (QUARTER) | Integer N=5003

Policy: 4/15-centered (CENTER) | Integer N=5003

Figure 1: Descent Graphs for $N = 5{,}003$ under all four deterministic policies: DOWN (top left), UP (top right), QUARTER (bottom left), CENTER (bottom right). Ellipses are internal nodes (label m = inbound coefficient mass); boxes are sinks (label w = final sink weight). Red edge labels indicate the ×2 Lemoine multiplier ($M = p + 2q$); unlabeled edges carry ×1. Not all ten sinks appear in every tree: sinks absent from a given decomposition receive $W_s = 0$, producing the structural zeros addressed by the pseudocount in Section 6.1.

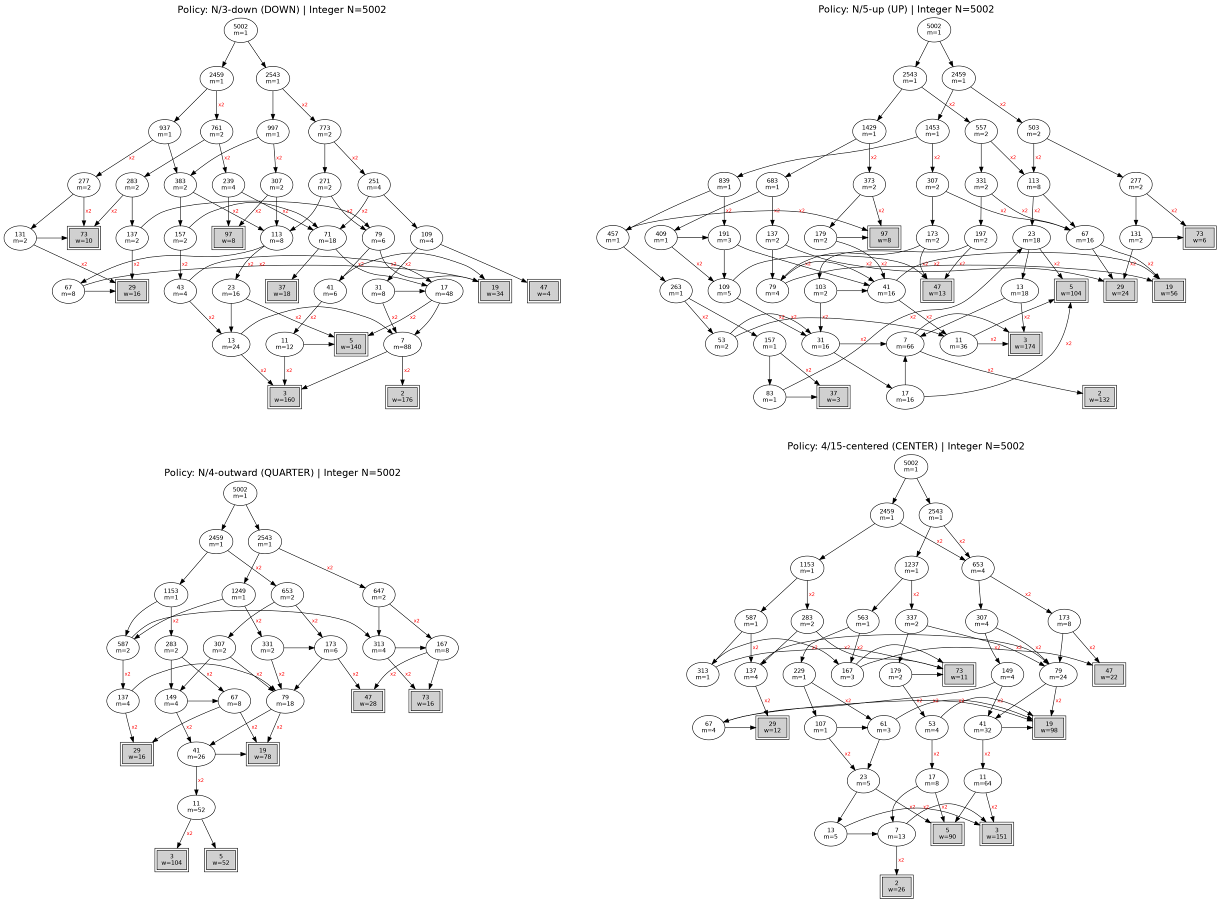


Figure 2: Descent Graphs for $N = 5{,}002$ (even), same four policies. The root undergoes Goldbach decomposition ($N = a + b$, both edges ×1); all subsequent odd nodes use Lemoine decomposition. Note the structural differences.

5 Attractors: the common sink set

5.1 The 10-sink terminal basis

One empirical finding is universal sink convergence. For every integer $N \geq 98$ in our sample, and across all four policies, the Descent Graph terminates at exactly ten non-consecutive primes:

$$\mathcal{S}_{10} = \{2, 3, 5, 19, 29, 37, 47, 59, 73, 97\}. \tag{1}$$

Several primes below 97 are absent from $\mathcal{S}_{10}$: for instance $7, 11, 13, 17, 23, 31, 41, 43, 53, 61, 67, 71, 79, 83, 89$. Their absence means that these primes, when encountered during descent, always admit a further constrained decomposition—they are never terminal. This fact can be verified by exhaustive enumeration:

Theorem 1 (Terminal basis determinism)**.** *A prime $N < 98$ is a sink (i.e., admits no constrained Lemoine witness $N = p + 2q$ with $q \in [\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ and p, q both prime) if and only if $N \in \mathcal{S}_{10}$.*

Proof. There are 25 primes below 98. For $N = 2$ the parity switch to Goldbach yields a trivially empty window. For $N = 3$ and $N = 5$, the window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ reduces to $\{1\}$, which is not prime. For each remaining prime N, Table 1 lists the window bounds; for non-sinks, it shows all valid witnesses. The seven non-trivial sinks $\{19, 29, 37, 47, 59, 73, 97\}$ each fail because every prime q in their window produces a composite $p = N - 2q$. For all 15 non-sink primes, at least one valid witness exists (rightmost column). ☐

Electing to iterate to exhaustion is not the only valid choice. Primes 2, 3, and 5 are sinks only because the Lemoine window collapses for very small integers; replacing them with the buffer primes $\{11, 13, 17, 23, 31\}$ that arise in their decomposition yields a 12-sink system:

$$\mathcal{S}_{12} = \{11, 13, 17, 19, 23, 29, 31, 37, 47, 59, 73, 97\}. \tag{2}$$

The qualitative geometry is preserved under this refinement, confirming that the phenomena we observe reflect the descent dynamics rather than the particular choice of terminal basis. The refinement mechanism and its algebraic structure are developed in Section 10.

6 Geometry of the arithmetic signatures

6.1 Compositional data and the ILR transform

The sink-weight vector $\mathbf{w}(N)$, once normalized, lies on the 10-simplex Δ^9. Simplicial data present well-known challenges for standard multivariate analysis: the constant-sum constraint induces spurious correlations, and Euclidean distances in raw proportions are distorted. Following Aitchison's framework for compositional data analysis [1], we apply the *isometric log-ratio* (ILR) transform, which maps Δ^9 isometrically into $\mathbb{R}^9$ via a Helmert orthonormal basis. After replacing structural zeros with a small pseudocount ($\varepsilon = 0.5$; see Appendix B for sensitivity analysis), closing rows to the simplex, computing centered log-ratios, and projecting onto the Helmert submatrix, each integer's signature becomes a point in 9-dimensional Euclidean space.

6.2 PCA and visualization

Principal component analysis of the ILR-transformed coordinates reveals that the variance structure is strikingly policy-dependent. Figure 3 displays the PC1–PC2 projection for all five policies, each

Table 1: Constrained Lemoine witnesses for all primes $N < 98$. For each prime, the window is $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$. A ✓ marks membership in $\mathcal{S}_{10}$ (no valid witness $N = p + 2q$ with p, q both prime and q in the window). For non-sinks, all valid witnesses are listed.

N	$\mathcal{S}_{10}$?	Window	Status
2	✓	empty	no witness
3	✓	$[1, 1]$	no witness
5	✓	$[1, 1]$	no witness
7		$[2, 2]$	$(p, q) = (3, 2)$
11		$[3, 3]$	$(5, 3)$
13		$[3, 4]$	$(7, 3)$
17		$[4, 5]$	$(7, 5)$
19	✓	$[4, 6]$	no witness
23		$[5, 7]$	$(13, 5)$
29	✓	$[6, 9]$	no witness
31		$[7, 10]$	$(17, 7)$
37	✓	$[8, 12]$	no witness
41		$[9, 13]$	$(19, 11)$
43		$[9, 14]$	$(17, 13)$
47	✓	$[10, 15]$	no witness
53		$[11, 17]$	$(31, 11)$; $(19, 17)$
59	✓	$[12, 19]$	no witness
61		$[13, 20]$	$(23, 19)$
67		$[14, 22]$	$(29, 19)$
71		$[15, 23]$	$(37, 17)$
73	✓	$[15, 24]$	no witness
79		$[16, 26]$	$(41, 19)$
83		$[17, 27]$	$(37, 23)$
89		$[18, 29]$	$(43, 23)$; $(31, 29)$
97	✓	$[20, 32]$	no witness

drawn from 20,000 integers sampled uniformly across four orders of magnitude (5×10^4–5×10^8) and colored by N mod 3. The five panels encode the paper's central claim at a glance: deterministic witness selection induces sharply differentiated geometry, while randomized selection produces none.

QUARTER (panel a) generates a sparse archipelago of near-discrete islands, with residue classes occupying separate clusters connected by thin filaments. CENTER (panel b) produces nested, concentric shell structures radiating outward from a dense core, with arcs organized by residue class. DOWN (panel c) forms a dominant spine along PC1 with periodic transversal striations—a signature of discrete lattice structure in the sink coefficients. UP (panel d) yields a diffuse cloud with modular neighborhoods; note the dramatically compressed PC1/PC2 scale (±0.6) compared to other policies. Finally, RANDOM (panel e) is an amorphous blob with fully interleaved residue classes—no geometric or modular structure emerges.

6.3 Multi-view analysis: center policy

We take the CENTER policy as a case study to illustrate the multi-scale and multi-dimensional structure of the embeddings. Adding magnitude ($\log_{10} N$) as a third axis (Figure 4) reveals that the concentric shells of Figure 3b are layered sheets that separate and evolve with scale. In the pure PC1–PC2–PC3 space (Figure 5), the filaments extend into three dimensions and residue classes occupy distinct sub-manifolds.

The mixed-magnitude projections above blend integers across four decades of N. Restricting to a fixed magnitude band sharpens the geometry considerably. Figure 6 shows 100,000 consecutive integers near 5×10^6. At fixed scale, CENTER separates into three tight, disjoint clusters—one per residue class—with almost no overlap. This effect is consistent with the dimensional contraction documented in Section 7.4: within a narrow magnitude band the slow variable ($\log N$) is approximately constant, and only the fast modular variables remain, producing near-discrete structure.

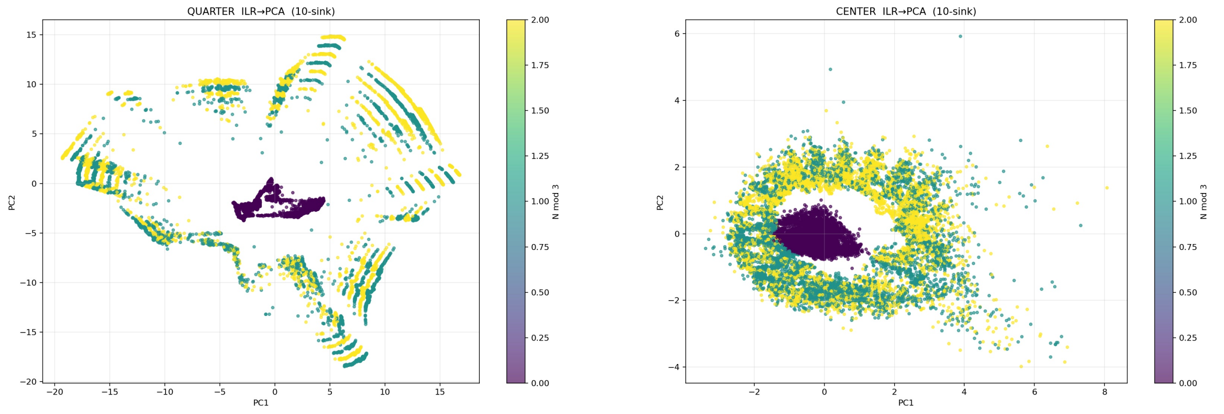
Remark 2 (Slow and fast variables). The embedding structure is governed by *slow variables* (scale $\log N$, window width, typical witness gap) that change gradually, and *fast variables* (N mod 3, N mod 6, early branch outcomes) that shift abruptly. Their interplay explains why adjacent integers have similar signatures while residue-class boundaries remain sharp. Fixing the slow variable (restricting to a narrow magnitude band) isolates the fast variables and yields the near-discrete separation seen in Figure 6. Changing only the witness-selection rule moves the dynamics along a spectrum from "quantized attractor" (Figure 3a) to "mixing flow" (Figure 3e).

7 Intrinsic dimension estimation

To characterize the embedding manifolds quantitatively, we estimate their intrinsic dimension using two complementary methods.

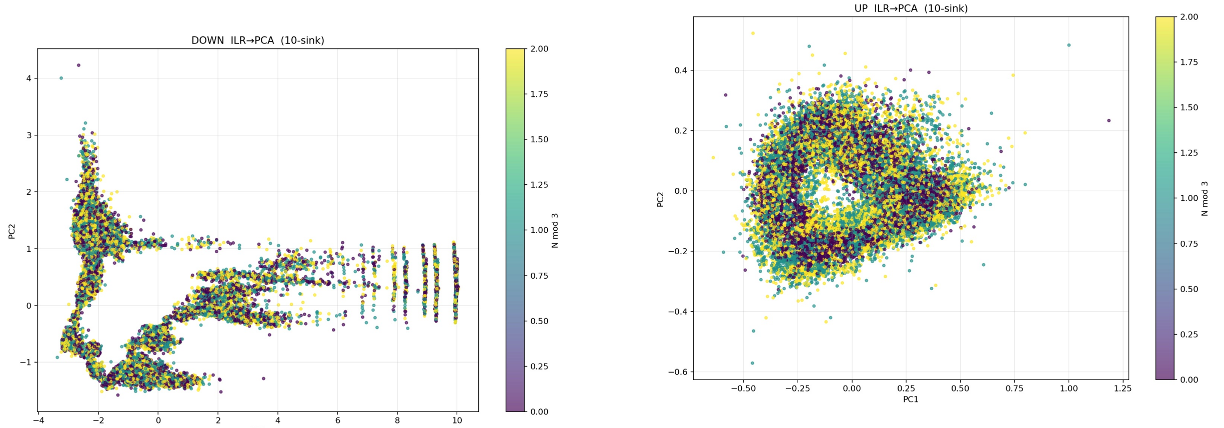
7.1 kNN-MLE (Levina–Bickel)

The Levina–Bickel estimator [4] uses the ratio of successive k-nearest-neighbor distances to produce a local dimension estimate at each point, which is then aggregated (we use a 2%-trimmed mean). We evaluate across $k \in \{10, 15, 20, 25, 30, 40, 50\}$. The kNN-MLE estimates show a characteristic downward trend with increasing k, reflecting the multi-scale nature of the point clouds: at small scales, local curvature and discreteness inflate the apparent dimension, while at larger scales the global manifold dimension emerges.



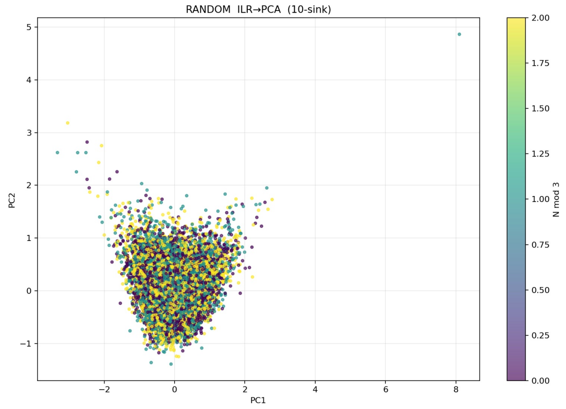
(a) QUARTER: archipelago of near-discrete islands / eccentric orbits.

(b) CENTER: nested shells with arc-and-filament connections.



(c) DOWN: spine with periodic transversal striations.

(d) UP: diffuse cloud with notable void and minimal residue-class organization. (note compressed scale).



(e) RANDOM: largely amorphous blob, no residue-class organization.

Figure 3: ILR→PCA embeddings (10-sink, PC1 vs. PC2) for all five policies, colored by N mod 3. Each panel shows 20,000 integers sampled uniformly across 5×10^4–5×10^8. Deterministic witness selection produces sharply differentiated geometry; randomized selection (panel e) destroys most or all apparent structure.

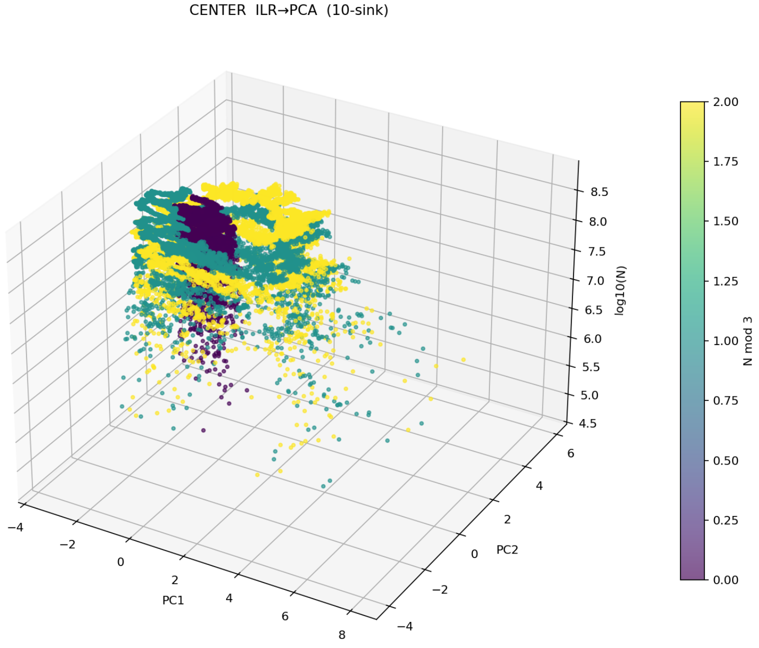


Figure 4: CENTER policy, 10-sink ILR→PCA: 3D projection (PC1, PC2, $\log_{10} N$), colored by N mod 3. The layered sheets reveal how structure evolves with scale.

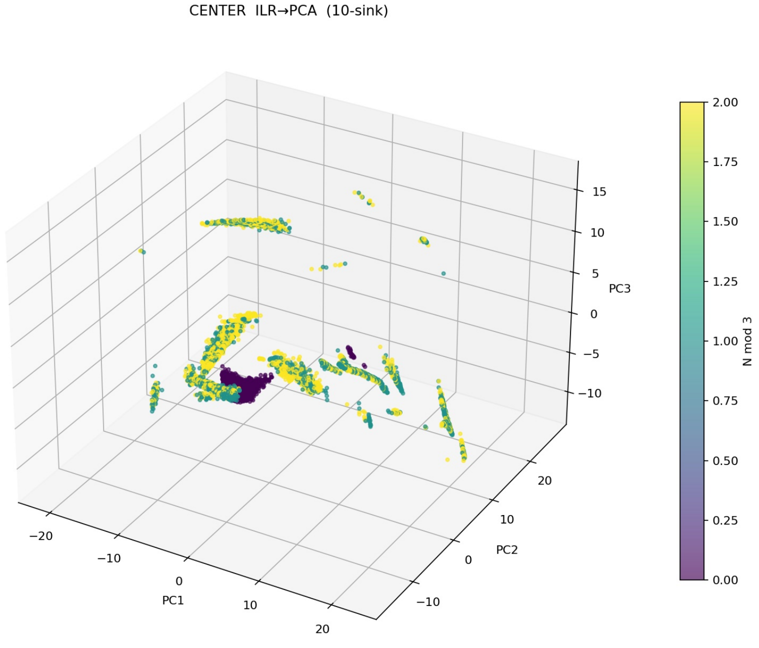


Figure 5: CENTER policy, 10-sink ILR→PCA: 3D projection (PC1, PC2, PC3), colored by N mod 3.

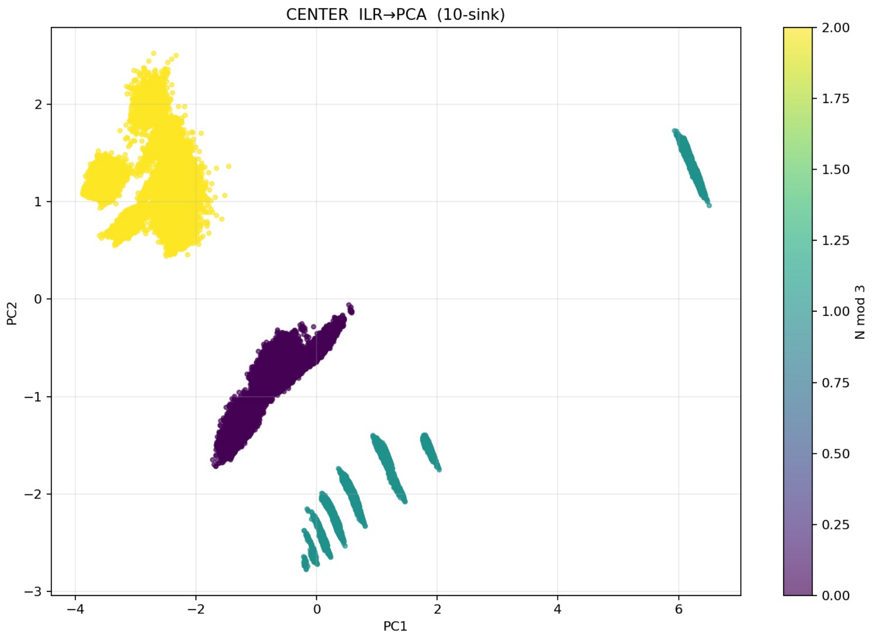


Figure 6: CENTER, 100K consecutive integers near 5×10^6: PC1 vs. PC2, colored by N mod 3. At fixed magnitude the three residue classes collapse into disjoint clusters.

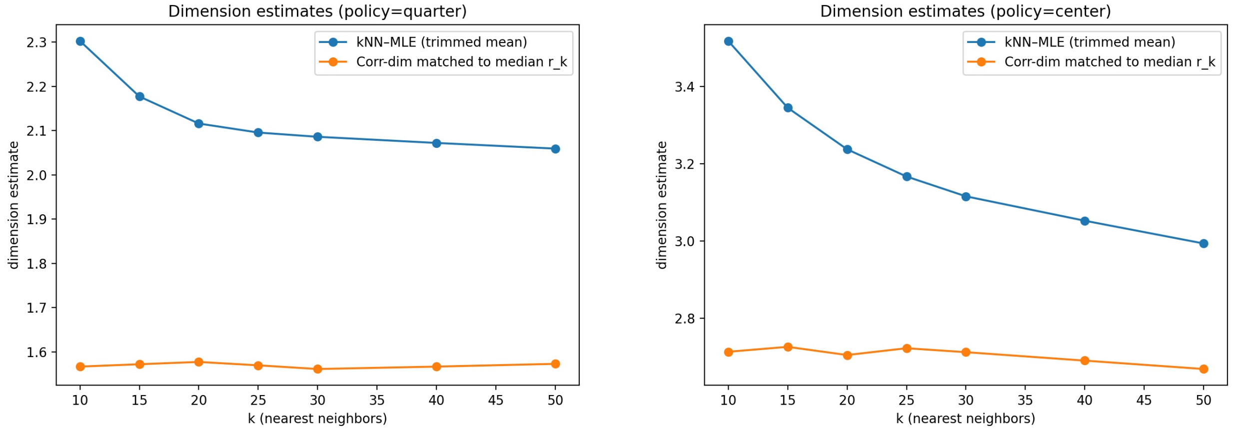
7.2 Correlation dimension (Grassberger–Procaccia)

We also estimate the correlation dimension via random-pair sampling [3] (1.5×10^6 pairs per policy). To ensure an apples-to-apples comparison with kNN-MLE, we match the correlation dimension fit window to the median kNN radius at each k, using a multiplicative band of width $\alpha = 2.0$ around the median. This "matched" correlation dimension is less sensitive to the choice of global radius grid and provides a second, independent dimension estimate at each spatial scale. The gap between kNN-MLE and correlation dimension is expected: kNN-MLE is a local estimator and tends to be inflated by multi-scale effects (curvature, local density variation), while correlation dimension is a global estimator that captures the dominant scaling regime. The true manifold dimension likely lies between the two, with correlation dimension providing a lower bound and kNN-MLE (at large k) an upper bound.

7.3 Results

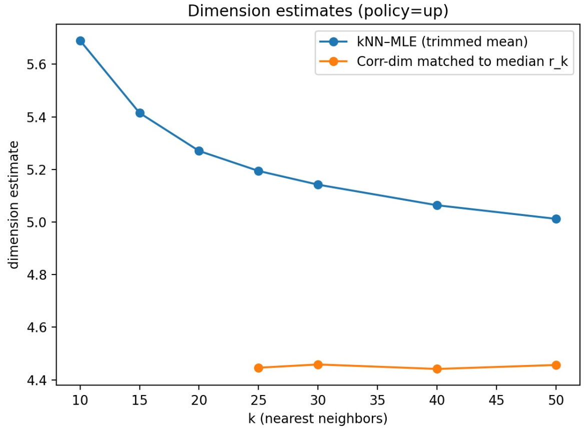
The results are consistent across methods and strikingly policy-dependent (Figure 7). The RANDOM null, with kNN-MLE $\hat{d} \approx 7.4$ (ambient dimension 9), confirms that the low dimensions observed under deterministic policies are genuine geometric constraints, not artifacts of the ILR transform or sample size.

Both estimators report a single global mean, but the manifold dimension need not be spatially uniform. The theory of Local Intrinsic Dimensionality (LID) [15] suggests that dimension can vary across sub-regions—for instance, the discrete islands of QUARTER may have lower internal dimension than the connecting bridges, and the residue-class filaments of CENTER may each inhabit a different-dimensional submanifold. Quantifying LID variation across these substructures is a natural direction for future work.



(a) QUARTER: kNN-MLE ≈ 2.3–2.06; matched corr-dim flat at ≈ 1.56.

(b) CENTER: kNN-MLE ≈ 3.45–3.0; matched corr-dim flat at ≈ 2.7.



(c) UP: kNN-MLE ≈ 5.7–5.0; matched corr-dim ≈ 4.45 (available only for $k \geq 25$).

Figure 7: Dimension estimates vs. k for three policies. Blue: kNN-MLE (trimmed mean); orange: correlation dimension matched to median kNN radius. The RANDOM policy (not shown) yields kNN-MLE ≈ 7.4 and correlation dimension ≈ 4.3.

7.4 Dimensional contraction across magnitude

Splitting the log-stratified sample into five magnitude bands ($[50\text{k}, 100\text{k}]$ through $[100\text{M}, 500\text{M}]$, each containing 4,000 integers), we find that deterministic policies exhibit a systematic decrease in $\hat{d}$ with increasing N (Figure 8).

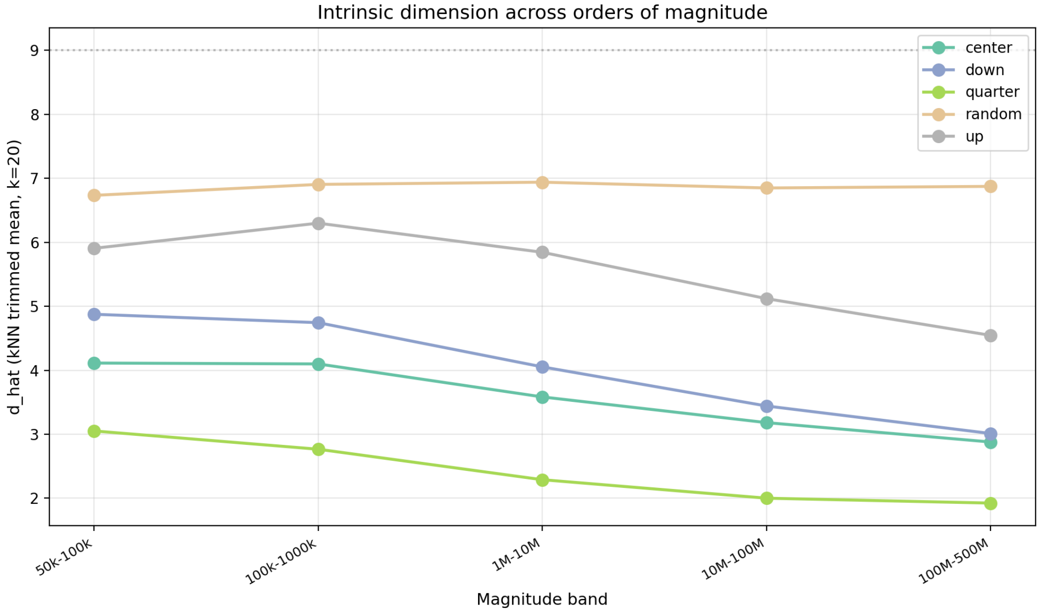


Figure 8: Intrinsic dimension (kNN-MLE, $k = 20$, trimmed mean) by magnitude band for all five policies. The policy ordering is preserved across all bands, but deterministic policies contract toward lower dimensions at larger N. The RANDOM null is flat ($\text{CV} = 0.011$).

The contraction is strongest for QUARTER ($\hat{d} = 3.05$ at $[50\text{k}, 100\text{k}]$ to 1.92 at $[100\text{M}, 500\text{M}]$) and weakest for UP (5.91 to 4.54). The RANDOM null shows no contraction (coefficient of variation $\text{CV} = 0.011$; range $[6.73, 6.94]$), confirming that the effect is policy-created, not a property of the witness pool or the ILR transform.

The mechanism is consistent with the modular arithmetic analysis of Section 8: at larger N, descent trees are deeper, and each additional level of the Lemoine recurrence accumulates further modular constraints on the sink coefficients. The rate of contraction varies by policy. With data spanning only 3.5 decades of N, we cannot distinguish whether the contraction is logarithmic, power-law, or some other form, nor whether it converges to a finite floor or continues indefinitely.

Remark 3. The within-band estimates are more faithful than the global $\hat{d}$ computed from the full mixed sample: mixing points from different magnitude bands, each living on a slightly different-dimensional submanifold, inflates the apparent global dimension. The global values in Table 2 and Figure 7 should be interpreted as upper bounds on the manifold dimension at any fixed scale.

8 Modular arithmetic of sink coefficients

The ILR transform discards the integer structure of the sink-weight vector $\mathbf{W}(N) = (W_2, W_3, \ldots, W_{97})$ by normalizing and taking logarithms. We now examine the raw integer coefficients modulo small primes, uncovering rigid arithmetic constraints that connect the witness-selection policy to the observed dimension hierarchy.

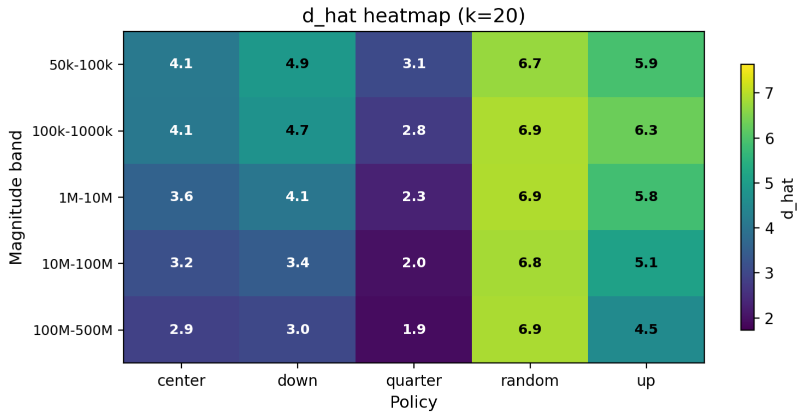


Figure 9: Intrinsic dimension heatmap (kNN-MLE, $k = 20$) by magnitude band and policy. The column-wise gradient in deterministic policies and the uniformity of RANDOM are visually immediate.

8.1 The reconstruction identity and modular constraints

The sink-weight vector satisfies the *reconstruction identity*

$$\sum_{s\in\mathcal{S}} s \cdot W_s(N) \;=\; N, \tag{3}$$

where $\mathcal{S} = \{2, 3, 5, 19, 29, 37, 47, 59, 73, 97\}$ is the 10-sink terminal basis. This is verified exactly for all 20,000 integers and all five policies (Experiment 1). Reducing (3) modulo a prime k yields

$$\sum_{s\in\mathcal{S}} (s \bmod k) \cdot (W_s \bmod k) \;\equiv\; N \pmod{k},$$

from which sinks s with $s \equiv 0 \pmod{k}$ "vanish" and impose no modular constraint. For $k = 2$: sink 2 vanishes; for $k = 3$: sink 3; for $k = 5$: sink 5. These vanishing patterns interact with the Lemoine recurrence to produce deterministic parity structure.

8.2 The parity skeleton

For any odd node $M = p + 2q$, the Lemoine recurrence gives

$$W_s(M) \;=\; W_s(p) + 2\,W_s(q) \;\equiv\; W_s(p) \pmod{2}. \tag{4}$$

The q-branch vanishes modulo 2 at every recursion level: the parity of W_s is determined entirely by the *p-chain*, the path that follows only p-children from M down to a terminal sink. The p-chain terminates at whichever sink it first reaches, contributing 1 to that sink and 0 to all others.

Proposition 1 (Parity skeleton)**.** *In the 10-sink system, sinks $\{2, 5, 59\}$ are p-*chain-unreachable*: no sequence of Lemoine p-children starting from any prime can terminate at these sinks. Consequently, $W_2(N)$, $W_5(N)$, and $W_{59}(N)$ are even for all integers $N \geq 98$, independent of policy.*

Proof sketch. Sink 2. In any Lemoine decomposition $N = p + 2q$ with N odd, $p = N - 2q$ is odd, so $p \geq 3$. The p-chain consists entirely of odd primes and can never reach 2.

Sink 5. The only prime N whose Lemoine window admits $p = 5$ is $M = 11$ (requiring $q = 3$). But 11 itself is unreachable as a p-child: for $M' = 11 + 2q$ to be prime with $q \in [\lceil M'/5 \rceil, \lfloor M'/3 \rfloor]$, the candidates are $q \in \{5, 7, 11\}$, giving $M' \in \{21, 25, 33\}$, none prime.

Sink 59. For $p = 59$ directly, the Lemoine constraint $q \in [\lceil M/5 \rceil, \lfloor M/3 \rfloor]$ with $M = 59 + 2q$ yields candidates $M \in \{101, 105, \ldots, 177\}$; checking each confirms that no such M is simultaneously prime and has q in-window.

All contributions to W_2, W_5, W_{59} therefore flow exclusively through q-branches, where the Lemoine coefficient of 2 ensures that each such contribution is even; the total is therefore even. For even roots $N = a + b$ in our tested ranges, both Goldbach children a, b are odd primes. $\square$

The remaining seven sinks *are* p-chain-reachable (e.g., $M = 41$ has $p = 19$; $M = 7$ has $p = 3$), so their coefficients can be odd. Figure 10 confirms: sinks $\{2, 5, 59\}$ show $\Pr[W_s \text{ even}] = 1.000$ across all policies and all residue classes $N \bmod 6$, while the other sinks show intermediate probabilities (60%–98%) depending on policy and residue class.

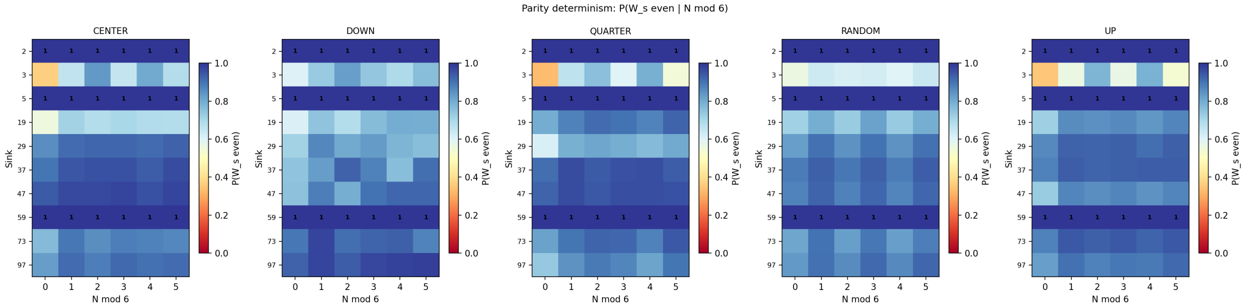


Figure 10: Parity determinism: $\Pr[W_s(N) \text{ even} \mid N \bmod 6]$ across all five policies and all sinks. Sinks $\{2, 5, 59\}$ are deterministically even (annotated "1"); the remaining sinks show policy-dependent parity distributions.

8.3 Mod-3 cancellation cascade

Modulo 3, the Lemoine recurrence becomes

$$W_s(M) \equiv W_s(p) - W_s(q) \pmod 3, \tag{5}$$

since $2 \equiv -1 \pmod 3$. Unlike the mod-2 case, *both* branches contribute—with opposite signs.

Under QUARTER, $q \approx N/4$ and $p = N - 2q \approx N/2$, so $p \approx 2q$. At the next recursion level, the p-child of p satisfies $p' \approx p/2 = N/4 \approx q$: the two subtrees rooted at p and q are highly correlated. In the mod-3 recurrence (5), subtracting correlated values produces a result biased toward 0. This *cancellation cascade* accumulates over multiple tree levels, driving large-sink coefficients toward divisibility by 3.

Figure 11 displays the conditional entropy $H(W_s \bmod 3 \mid N \bmod 6)$ for each policy and sink. The QUARTER panel shows dramatic entropy collapse at sinks $\{29, 37, 47, 59, 73, 97\}$—all sinks reached deep in the tree, where the cascade has had multiple levels to accumulate. The RANDOM panel is uniformly high entropy ($H \approx \log_2 3 \approx 1.58$), confirming that the mod-3 structure is entirely policy-created.

The cascade is policy-dependent because the subtree correlation depends on the ratio p/q at each level: QUARTER ($p/q \approx 2$) has the strongest correlation, DOWN ($p \approx q$, moderate correlation) is intermediate, and UP ($p \approx 3q$, weak correlation) shows almost no mod-3 structure. This gradient in

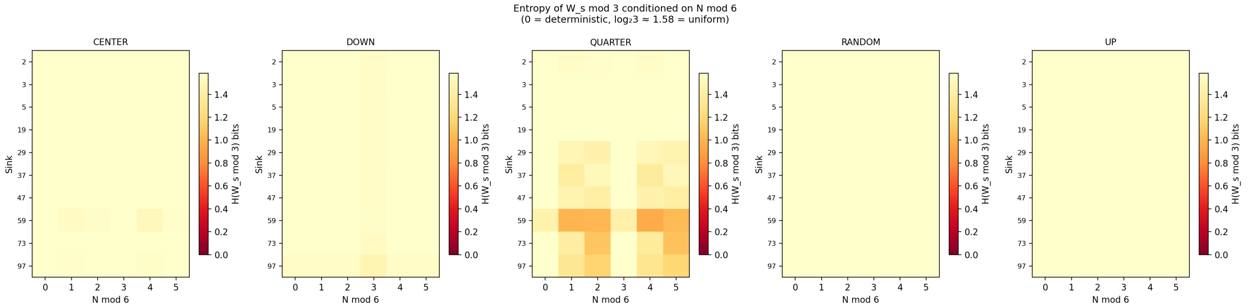


Figure 11: Conditional entropy $H(W_s \bmod 3 \mid N \bmod 6)$ by policy and sink (0 = deterministic, $\log_2 3 \approx 1.58$ = uniform). QUARTER shows extreme entropy collapse at the large sinks, RANDOM is uniformly maximal.

modular concentration maps directly onto the dimension hierarchy: stronger mod-3 concentration → fewer effective degrees of freedom in the coefficient vector → lower manifold dimension.

Remark 4. The cross-sink GCD structure further confirms the modular linkage. Under QUARTER, the mean $\gcd(W_2, W_{97})$ across all integers is $\approx 1.4 \times 10^6$, compared with ≈ 39 under RANDOM and ≈ 11 under UP. These enormous shared factors are a direct consequence of the mod-3 (and higher-order) concentration.

9 Persistent homology

9.1 Method

To probe the topological structure of the ILR embeddings beyond dimension, we compute persistent homology through H_2 using the Vietoris–Rips filtration (implemented via Ripser [6]). For each policy, we subsample 800 points from the ILR-transformed coordinates and run 10 replicates with different random seeds to assess stability. We report the count of finite-lifetime features, average lifetime, and maximum lifetime for each homological degree and policy.

9.2 Results

The persistence diagrams reveal a clean dimensional hierarchy across policies. Table 2 summarizes the H_1 and H_2 statistics (mean ± std over 10 replicates, 12-sink ILR, 800-point subsamples).

Table 2: Persistent homology summary (12-sink, ILR, H_2). Statistics are mean ± std over 10 subsampling replicates of 800 points.

Policy	$\hat{d}_{\text{corr}}$	H_1 count	H_1 avg life	H_1 max life	H_2 count	H_2 avg life	H_2 max life
QUARTER	1.56	56 ± 4	0.70 ± 0.08	16.8 ± 1.7	1.3 ± 0.7	0.04	—
DOWN	≈ 2	155 ± 14	0.039 ± 0.001	0.82 ± 0.19	12 ± 3	0.012 ± 0.004	0.04 ± 0.02
CENTER	2.7	293 ± 12	0.060 ± 0.002	1.04 ± 0.04	76 ± 6	0.023 ± 0.004	0.14 ± 0.04
UP	4.45	440 ± 14	0.012 ± 0.001	0.15 ± 0.01	118 ± 14	0.005 ± 0.0004	0.03 ± 0.01

9.3 Interpretation

quarter has the fewest H_1 features (56 per replicate) but they are enormous—average lifetime 0.70, maximum ≈ 17. H_2 is nearly absent (0–2 features per replicate). This is the topological signature of an *archipelago*: widely separated clusters connected by long bridges. Each H_1 feature corresponds to a cycle through multiple islands, and the huge lifetimes mean these cycles persist across a wide range of filtration scales. The near-zero H_2 count confirms there are no 2-dimensional voids—the space between islands is genuinely empty. Topologically, this is a graph (1-complex) at the relevant scale, consistent with the intrinsic dimension of ≈ 1.6.

down has moderate H_1 (155 features, short-lived) and almost no H_2 (10–15 features with tiny lifetimes). This matches the near-one-dimensional spine: the few long-lived H_1 features (max ≈ 0.8) likely correspond to rare points where the spine curves back on itself or where residue-class filaments come close together in ILR space. The near-absence of H_2 is exactly what one expects from intrinsic dimension ≈ 2.

center has the richest topological structure: abundant H_1 (≈ 293 features, stable across replicates) and robust H_2 (≈ 76 features, maximum lifetime ≈ 0.14). The H_1 features correspond to the closed loops visible in the PC1–PC2 arc-and-filament structure. The H_2 features indicate genuine 2-dimensional voids—consistent with the layered sheet structure visible in the 3D projections, where sheets at different $\log_{10} N$ levels enclose cavities. CENTER at dimension ≈ 3 sits in the sweet spot where H_2 features are prominent but not yet washed out.

up has the most H_1 features by count (≈ 440) but they are all tiny (avg 0.012, max 0.15). H_2 is also abundant (118 features) but equally short-lived (avg 0.005). This is the signature of a relatively high-dimensional, well-mixed cloud: topological features form and die quickly as the filtration radius grows because the manifold is thick and everything gets filled in. The abundance of short-lived features is also a known artifact of Vietoris–Rips filtrations in high ambient dimensions, where sampling noise alone generates many ephemeral cycles. The topology is therefore less well-articulated, not because the manifold is simple, but because it is too high-dimensional for low-degree homology to persist—in contrast to the stable voids of CENTER, which arise from more well-defined geometric structure.

random extends the hierarchy to its natural upper bound. On a separate 10-sink, log-stratified sample (1,000-point subsamples, 10 replicates), RANDOM produces the most H_1 features (≈ 839) and by far the most H_2 (≈ 416), but all are short-lived (avg H_1 lifetime 0.039, avg H_2 lifetime 0.017). This is the topological null: the point cloud nearly fills the simplex, so cycles and voids form and collapse at every scale with no persistent structure.

Across all five policies, we observe a clean dimensional hierarchy: as intrinsic dimension increases from QUARTER → DOWN → CENTER → UP → RANDOM, the topology transitions from "sparse, long-lived H_1 with no H_2" (archipelago) through "abundant H_1 and emerging H_2" (structured manifold with loops and voids) to "abundant but short-lived features across all degrees" (high-dimensional blob). This is precisely what the theory of persistent homology on Riemannian manifolds predicts: features at homological degree d are prominent when the intrinsic dimension is close to $d+1$, and get washed out when the dimension is much larger. The RANDOM null at the top of this hierarchy confirms that the topological structure of the deterministic policies is created by the selection rule, not inherited from the witness pool.

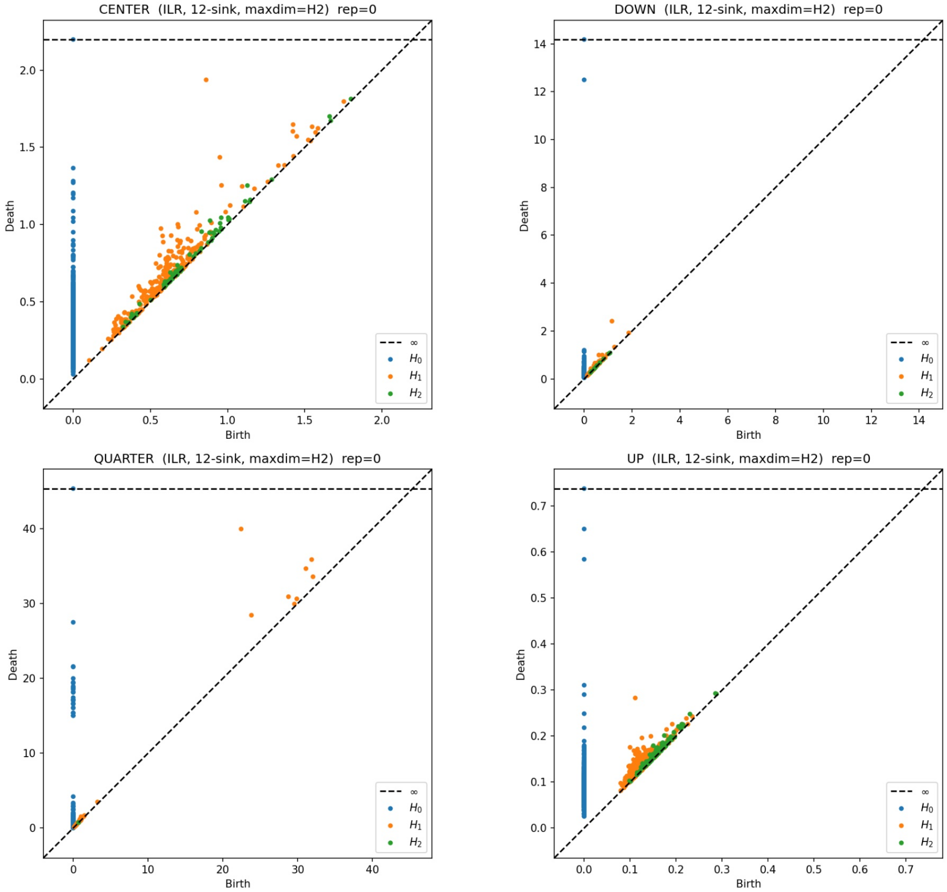
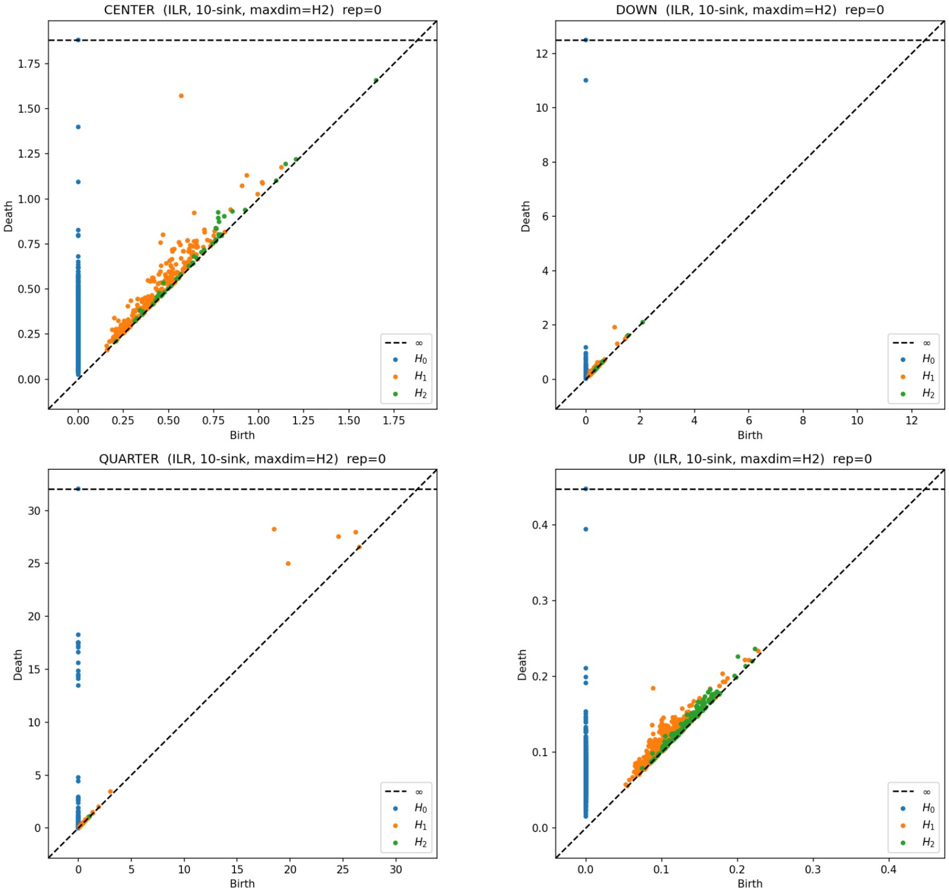


Figure 12: Persistence diagrams (12-sink, ILR, H_2): CENTER (top left), DOWN (top right), QUARTER (bottom left), UP (bottom right). Points far from the diagonal have long lifetimes. Note the dramatically different axis scales across policies.



CENTER (ILR, 10-sink, maxdim=H2) rep=0

DOWN (ILR, 10-sink, maxdim=H2) rep=0

QUARTER (ILR, 10-sink, maxdim=H2) rep=0

UP (ILR, 10-sink, maxdim=H2) rep=0

Figure 13: Persistence diagrams (10-sink, ILR, H_2): same layout. The qualitative pattern is preserved under terminal basis coarsening, confirming topological robustness.

10 Terminal basis refinement and the poset of bases

The preceding section used the 10-sink and 12-sink terminal bases side by side, confirming that the qualitative geometry and topological structure are all preserved under the change of basis. We now develop the algebraic framework that explains *why* this must be so: both bases are cross-sections of the same canonical descent tree, and the relationship between any two admissible bases is governed by a linear coarsening map.

10.1 The refinement mechanism

One can "resolve" a sink s by removing it from the terminal set and adding the parent set from which it was germinated under the constrained Lemoine window. Applying this to primes 2, 3, and 5 (which are sinks only because the window collapses for very small integers) and replacing them with the buffer primes $\{11, 13, 17, 23, 31\}$ produces the 12-sink system $\mathcal{S}_{12}$ (2). Conversely, the 10-sink set is a *coarsening* of the 12-sink set: sinks $11, 13, 17, 23, 31$ are absorbed back into the tree and their mass redistributed to $2, 3, 5$. Indeed, the 12-sink PCA projections are often somewhat *sharper* than their 10-sink counterparts: sinks 2, 3, and 5 carry the largest mass but also the most local variability, so their removal distributes the embedding's variance across a wider basis.

10.2 The poset of terminal bases

The refinement operation is not a one-off construction: it can be iterated, and the resulting family of terminal bases carries a rigid algebraic structure. The key insight is that the full descent tree $G_\pi(N)$ is the canonical object; different sink sets correspond to different cross-sections of the same tree, and stopping rules can be incorporated into the algorithm.

Definition 3 (Admissible terminal basis)**.** Fix a policy π and a search window (in our case, $q \in [\lceil M/5 \rceil, \lfloor M/3 \rfloor]$). A finite set of primes $\mathcal{S}$ is an *admissible terminal basis* if every integer $N > \max(\mathcal{S})$ eventually reaches a prime in $\mathcal{S}$ under iterated π-descent within the given window.

Window-irreducibility and the minimal basis. Theorem 1 establishes that each of the ten primes in $\mathcal{S}_{10}$ (1) admits *no* constrained Lemoine witness within the window $[\lceil M/5 \rceil, \lfloor M/3 \rfloor]$. All ten are equally irreducible under the present window constraints. This irreducibility is a property of the window, not of the primes alone: a wider window could render some of these primes decomposable and would yield a different minimal basis. Under the present window, $\mathcal{S}_{10}$ is the unique minimal admissible terminal basis: removing any sink s would leave the descent stuck at s, since s can neither decompose nor be bypassed.

Refinement by shielding. Since no sink in $\mathcal{S}_{10}$ can be resolved by further decomposition within the constrained window, the refinement operation proceeds by *shielding*: adding new, larger primes that intercept descent paths *before* they reach certain $\mathcal{S}_{10}$-sinks. The 12-sink system $\mathcal{S}_{12}$ (2) replaces $\{2, 3, 5\}$ with the buffer primes $\{11, 13, 17, 23, 31\}$, which lie above $\{2, 3, 5\}$ in every descent tree and intercept all paths that would otherwise reach them. In the $\mathcal{S}_{12}$-descent, the primes $2, 3, 5$ are simply never encountered.

The creation of the 12-sink set is somewhat arbitrary. It may be generalized. Write $\mathcal{S} \preceq \mathcal{S}'$ if $\mathcal{S}'$ refines $\mathcal{S}$ (i.e., $\mathcal{S}'$ interposes additional sinks that shield some $\mathcal{S}$-sinks, so that the $\mathcal{S}'$-descent truncates the tree earlier than the $\mathcal{S}$-descent). Under $\preceq$, the admissible bases form a partially ordered set with unique minimum $\mathcal{S}_{10}$:

$$\mathcal{S}_{10} \preceq \mathcal{S}_{12} \preceq \mathcal{S}_{14} \preceq \cdots$$

where each $\mathcal{S}_{2k}$ is obtained by further resolving some $\mathcal{S}_{2(k-1)}$-sinks via additional buffer primes. Each step adds buffer primes that shield some subset of the previously reachable sinks. The chain extends without limit: at each stage, the new buffer sinks are primes whose own descent trees (under the current basis) terminate at the current sinks, so making them terminal is consistent.

The tree as canonical object. The descent tree $G_\pi(N)$—run under policy π with no terminal basis, i.e., continuing until every leaf is window-irreducible—is the fundamental object. An admissible basis $\mathcal{S}$ defines a *truncation* of $G_\pi(N)$: halt the descent at each node belonging to $\mathcal{S}$, and record the resulting sink weights. A coarser basis $\mathcal{S}$ truncates later (deeper in the tree); a finer basis $\mathcal{S}'$ truncates earlier (higher in the tree). The weight vector $\mathbf{W}_{\mathcal{S}}(N)$ is thus a projection of the full tree onto the cross-section defined by $\mathcal{S}$.

Since the reconstruction identity (3) recovers N from $\mathbf{W}_{\mathcal{S}}(N)$, and N together with π uniquely determines the full tree $G_\pi(N)$, the weight vector at *any* admissible basis determines the full tree and hence the weight vector at every other basis:

$$\mathbf{W}_{\mathcal{S}}(N) \xrightarrow[\text{identity}]{\text{reconstruction}} N \xrightarrow{\pi\text{-descent}} G_\pi(N) \xrightarrow[\text{at } \mathcal{S}']{\text{truncate}} \mathbf{W}_{\mathcal{S}'}(N). \tag{6}$$

Every arrow is deterministic and constructive.

Linearity of coarsening. Although the full chain (6) passes through the tree, the coarsening direction admits a shortcut. Every fine sink $t \in \mathcal{S}'$ either belongs to $\mathcal{S}$ already or lies above some $\mathcal{S}$-sinks in the tree; in either case, its own descent against $\mathcal{S}$ produces a fixed weight vector $\mathbf{W}_{\mathcal{S}}(t)$ independent of the root N (because t is a fixed integer and the policy is deterministic).

Lemma 1 (Linear coarsening)**.** *Fix a policy π and admissible bases $\mathcal{S} \preceq \mathcal{S}'$. For each fine sink $t \in \mathcal{S}'$, let $\mathbf{W}_{\mathcal{S}}(t) \in \mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$ denote the weight vector of t under the $\mathcal{S}$-descent. (If $t \in \mathcal{S}$, this is the standard basis vector e_t.) Then for every $N > \max(\mathcal{S}')$,*

$$W_s^{\mathcal{S}}(N) \;=\; \sum_{t \in \mathcal{S}'} W_t^{\mathcal{S}'}(N) \;\cdot\; W_s^{\mathcal{S}}(t), \qquad \forall\, s \in \mathcal{S}. \tag{7}$$

Proof. Since $\mathcal{S}'$ truncates the tree earlier than $\mathcal{S}$, every unit of weighted mass flowing from the root N toward an $\mathcal{S}$-sink s is first intercepted at exactly one fine sink $t \in \mathcal{S}'$—the first $\mathcal{S}'$-member encountered along the path. (This interception is unique because once the path reaches a $\mathcal{S}'$-sink the $\mathcal{S}'$-descent stops; the subsequent descent from t to $\mathcal{S}$-sinks is a separate, fixed computation.) The total mass intercepted at t is $W_t^{\mathcal{S}'}(N)$. From t, the descent continues under policy π against the coarser basis $\mathcal{S}$, distributing a fixed fraction $W_s^{\mathcal{S}}(t)$ to each coarse sink s—this quantity depends only on t, π, and $\mathcal{S}$, not on the root N. Summing over all fine sinks yields the result. $\square$

In matrix form, let $C = C_\pi(\mathcal{S}', \mathcal{S})$ be the $|\mathcal{S}| \times |\mathcal{S}'|$ *coarsening matrix* whose (s,t)-entry is $W_s^{\mathcal{S}}(t)$. Then

$$\mathbf{W}_{\mathcal{S}}(N) \;=\; C\,\mathbf{W}_{\mathcal{S}'}(N). \tag{8}$$

Columns corresponding to shared sinks $t \in \mathcal{S} \cap \mathcal{S}'$ are standard basis vectors; columns corresponding to buffer sinks $t \in \mathcal{S}' \setminus \mathcal{S}$ encode how their mass redistributes to the coarse sinks when the descent continues past them.

For the $\mathcal{S}_{12} \to \mathcal{S}_{10}$ coarsening, the matrix is displayed in Table 3. It is policy-invariant: the five buffer sinks $\{11, 13, 17, 23, 31\}$ are small enough that their Lemoine windows each admit only a single witness, so all policies produce the same decomposition.

Table 3: Coarsening matrix $C(\mathcal{S}_{12}, \mathcal{S}_{10})$. Each column is the $\mathcal{S}_{10}$-weight vector of the corresponding $\mathcal{S}_{12}$-sink. The matrix is policy-invariant: the buffer sinks are small enough that their Lemoine windows admit only one witness each.

$s \backslash t$	**11**	**13**	**17**	19	**23**	29	**31**	37	47	59	73	97
2	·	2	2	·	2	·	6	·	·	·	·	·
3	2	3	1	·	3	·	3	·	·	·	·	·
5	1	·	2	·	2	·	2	·	·	·	·	·
19	·	·	·	1	·	·	·	·	·	·	·	·
29	·	·	·	·	·	1	·	·	·	·	·	·
37	·	·	·	·	·	·	·	1	·	·	·	·
47	·	·	·	·	·	·	·	·	1	·	·	·
59	·	·	·	·	·	·	·	·	·	1	·	·
73	·	·	·	·	·	·	·	·	·	·	1	·
97	·	·	·	·	·	·	·	·	·	·	·	1

Theorem 2 (Basis equivalence)**.** *Fix a policy π and admissible bases $\mathcal{S} \preceq \mathcal{S}'$.*

(i) (Injectivity.) *For every admissible basis $\mathcal{S}$, the map $N \mapsto \mathbf{W}_{\mathcal{S}}(N)$ is injective on $\{N \in \mathbb{N} : N > \max(\mathcal{S})\}$.*

(ii) (Full recoverability.) *The weight vector $\mathbf{W}_{\mathcal{S}}(N)$ determines the full descent tree $G_\pi(N)$ and hence the weight vector $\mathbf{W}_{\mathcal{S}'}(N)$ at every other admissible basis $\mathcal{S}'$, via the chain* (6)*.*

(iii) (Linear coarsening.) *The coarsening direction admits the matrix shortcut $\mathbf{W}_{\mathcal{S}}(N) = C\, \mathbf{W}_{\mathcal{S}'}(N)$.*

(iv) (Bijection.) *The change-of-basis map $\mathbf{W}_{\mathcal{S}}(N) \leftrightarrow \mathbf{W}_{\mathcal{S}'}(N)$ is a bijection between the respective image sets. Both directions are deterministic and constructive.*

Proof. (i) The reconstruction identity provides a left inverse: $\mathbf{W}_{\mathcal{S}}(N) \mapsto \sum_{s \in \mathcal{S}} s \cdot W_s(N) = N$.

(ii) From (i), $\mathbf{W}_{\mathcal{S}}(N)$ determines N. The policy π and the window constraints determine the descent tree $G_\pi(N)$ uniquely. Any admissible basis $\mathcal{S}'$ defines a truncation of this tree, from which $\mathbf{W}_{\mathcal{S}'}(N)$ is read off directly.

(iii) Lemma 1.

(iv) Combine (i) and (ii): the composite $\mathbf{W}_{\mathcal{S}}(N) \to N \to G_\pi(N) \to \mathbf{W}_{\mathcal{S}'}(N)$ is a bijection on image sets, with the reverse direction given by the same chain starting from $\mathcal{S}'$. In the coarsening direction, the matrix C provides a more efficient route. □

Remark 5 (Asymmetry of linearity). Coarsening is linear (a matrix multiplication) because the subtree below each fine sink is independent of N. Refinement is not: a coarse sink $s \in \mathcal{S} \setminus \mathcal{S}'$ is window-irreducible and never appears in the $\mathcal{S}'$-descent, so it has no $\mathcal{S}'$-weight vector. The $\mathcal{S}$-weight vector records how much mass reached s but not *how* it got there—and the "how" determines where the fine-basis buffer sinks would intercept the paths. Recovering this requires reconstructing the tree. The bijection is nonetheless fully constructive: the reconstruction identity and a single π-descent suffice.

Corollary 1 (Poset structure)**.** *The admissible terminal bases under π (for a fixed search window), ordered by refinement, form a countable poset with:*

(i) *a unique minimal element $\mathcal{S}_{10}$ (all of whose members are window-irreducible under the constrained Lemoine window);*

(ii) *at every pair* $\mathcal{S} \preceq \mathcal{S}'$*, a constructive bijection between weight-vector image sets that preserves* N*—equivalently, all admissible bases encode the same arithmetic information, since each determines the full tree;*

(iii) *composability of coarsening:* $C_\pi(\mathcal{S}'', \mathcal{S}) = C_\pi(\mathcal{S}', \mathcal{S})\, C_\pi(\mathcal{S}'', \mathcal{S}')$ *for* $\mathcal{S} \preceq \mathcal{S}' \preceq \mathcal{S}''$*.*

Proof. Part (i) follows from the window-irreducibility of all ten sinks (Theorem 1). Part (ii) is Theorem 2(iv). Part (iii) is immediate from path factoring: coarsening from $\mathcal{S}''$ to $\mathcal{S}$ through the intermediate basis $\mathcal{S}'$ factors each descent path into two segments, yielding the matrix product. ☐

Corollary 2 (Compositional amalgamation)**.** *The coarsening map* C *induces a subcompositional amalgamation in the sense of Aitchison [1] on the row-normalized compositions: it merges fine-basis components whose mass flows to a common coarse sink. Under an ILR basis chosen compatible with the amalgamation partition, this amalgamation admits an orthogonal-projection interpretation; under a generic Helmert basis (such as the one used in Section 6.1), it remains a valid subcompositional operation but the projection is oblique. The refinement direction (adding resolution) is not a linear operation on compositions; it requires reconstructing the integer and re-truncating the tree.*

Geometric implications. Theorem 2 guarantees that every admissible terminal basis encodes exactly the same arithmetic information about N: any weight vector determines the full tree, from which any other weight vector is recoverable. The *geometry* of the ILR manifold, however, need not be invariant across bases. Coarsening projects the fine manifold onto a lower-dimensional ambient space via the matrix C; refinement embeds the same information into a higher-dimensional space (though via a nonlinear route). Three regimes are possible for the intrinsic dimension d as the number of sinks k increases along the refinement chain: (a) d/k approximately constant; (b) d approximately constant (extra coordinates are redundant); (c) d grows sublinearly. The existing estimates at $k = 10$ and $k = 12$ (Section 7.4) provide two points on this curve per policy. Extending the computation to $k = 14$ or beyond would distinguish these regimes.

Window dependence. All of the above holds for the constrained window $[\lceil M/5 \rceil, \lfloor M/3 \rfloor]$. Under a different window, some members of $\mathcal{S}_{10}$ might become decomposable, shrinking (or enlarging) the minimal basis and altering the entire poset. The poset structure is a joint invariant of the policy, the window, and the prime landscape.

Upper bound on the poset. For a fixed tree $G_\pi(N)$, the number of distinct primes at each depth first grows (as the tree fans out from the root) and then contracts (as paths converge to the finite sink set at the leaves). Any admissible basis corresponds to a cross-section of this tree; the width of the cross-section reflects the sink count. For a *universal* basis—one that works simultaneously for all $N > \max(\mathcal{S})$—the buffer primes must intercept paths in *every* tree. But primes appearing near depth 1 in $G_\pi(N)$ are of order $N/2$ and appear near the top of only a few trees: those of integers close to N. As the buffer primes in a refinement chain grow larger, each one's "catchment" thins, and eventually no finite set of new buffer primes can universally intercept all remaining paths to the shielded sinks. This suggests the refinement chain terminates: the poset has not only a unique minimum ($\mathcal{S}_{10}$) but plausibly also a maximal admissible basis (or at least a finite upper bound on sink count), whose size would be a policy-dependent invariant. Determining this upper bound—or even whether it is sharp or merely asymptotic—is an open problem closely related to the ascending construction discussed in Question 3.

11 Memory structure

The descent dynamics exhibit a duality in their memory structure. *Vertically*—along the depth of the tree—memory is short: the first one or two decomposition steps capture most of the information about the terminal sink vector, with predictability strongest for DOWN and weakest for QUARTER. *Horizontally*—across consecutive integers—deterministic policies produce strong serial dependence, but the origin is ultimately simple: at each depth d, each policy begins its witness search at a fixed anchor point in a window that shifts by only $\sim 1/3$ to $\sim 1/5$ per unit increment of node M, and because valid witnesses are abundant, those selected for integer N throughout the decomposition process have an elevated probability of selection across many consecutive integers. The result is a tightly constrained trajectory that slowly evolves with N. The random policy, which selects a fresh anchor point in the $(M/5, M/3)$ window at each step, scrambles the trajectories entirely and yields white noise ($H = 0.49$).

This section outlines, quantifies, and explains the effect with three diagnostic layers—residue-class entropy, spectral structure, and Hurst exponents. The evidence for high long-term memory is persuasive. Equally persuasive is the evidence that this memory is a function of the deterministic selection rule, not the raw distribution of witnesses.

11.1 Entropy and spectral stratification

Treating the sink-weight sequence as a time series indexed by consecutive integers reveals dependence dominated by arithmetic periodicity. Shannon entropy $H(\mathbf{w}(N))$ varies systematically with $N \bmod 6$, with the governing modulus depending on policy: DOWN stratifies by parity (a gap of 0.15 bits between even and odd residues, within-class SD < 0.016); QUARTER stratifies by $N \bmod 3$ (1.82 to 2.21 bits, with mod-6 adding nothing); UP is nearly flat (total range 0.04 bits); and CENTER interpolates, with mod-3 dominant and mod-6 secondary. These patterns reflect each policy's mechanical sensitivity to the root's residue class: DOWN's witness availability depends on parity, QUARTER's on mod-3 density near $N/4$, and UP's on a region where prime density is least residue-sensitive.

The autocorrelation function confirms this picture. The raw ACF shows strong lag-6 periodicity ($\rho_6 \approx +0.95$ for DOWN); after subtracting the mean of each $N \bmod 6$ class, the lag-1 ACF drops to near zero for all policies, confirming that the periodic component accounts for almost all short-lag structure. However, residual autocorrelation persists at longer lags: at lag 30, the detrended ACF is +0.46 (DOWN) and +0.54 (UP), exceeding the 95% threshold (± 0.006 for $T = 100{,}000$) by orders of magnitude. Under the random policy, the detrended ACF at all lags is indistinguishable from zero.

11.2 Serial dependence: Hurst exponents

To quantify the residual persistence, we estimate Hurst exponents via DFA(1) [16] on 100,000 consecutive integers ($N = 5{,}000{,}001$ to $5{,}100{,}000$).

Main finding. After removing mod-6 periodicity and secular drift (a running-median detrend within each residue class), the ILR series under all four deterministic policies exhibit strong persistence ($H \approx 0.76$–0.86), confirmed at $Z = 18$–60 above a within-class shuffle null. Under RANDOM, $H = 0.49$ across all coordinates and detrending stages ($Z = 1.12$, $p = 0.12$). Multifractal DFA ($q \in [-5, 5]$) yields singularity width $\Delta\alpha < 0.06$—monofractal scaling consistent with a single mechanism (Table 4; Figure 14).

Mechanism. The persistence has a straightforward origin. Each deterministic policy anchors its witness search at a fixed fraction of M (e.g., $M/3$ for DOWN, $M/4$ for QUARTER). As N increments by 1, the Lemoine window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$ shifts by a fraction of a unit, but the gaps between valid witnesses are much larger than this shift. The nearest witness to the anchor therefore often remains the same for many consecutive integers, producing plateaus that generate $H > 0.5$. The random policy selects independently at each step, destroying the plateaus and yielding $H = 0.49$—confirming that the persistence is a property of the interaction between the selection rule and the prime landscape, not of the prime distribution alone.

Table 4: Hurst exponents via DFA(1), averaged over the first 6 ILR coordinates, under three detrending stages. "Full" denotes mod-6 plus secular trend removal. Z-scores are against a within-class shuffle null (50 permutations). The RANDOM policy serves as a null test for the deterministic selection mechanism.

Policy	Mean H (raw)	Mean H (mod-6)	Mean H (full)	Null H	Z-score
CENTER	0.58	0.80	**0.82**	0.62	18.7
DOWN	0.76	0.92	**0.79**	0.48	60.3
QUARTER	0.35	1.00	**0.86**	0.49	54.1
UP	0.81	0.88	**0.76**	0.48	55.1
RANDOM	0.50	0.50	0.49	0.48	1.1

11.3 Arithmetic substructure: primes and twin primes

As a probe of the embedding's sensitivity to number-theoretic structure, we compare primes versus composites and twin prime pairs versus generic lag-2 pairs within the 100,000-integer block (6,458 primes, 513 twin pairs). Under DOWN, primes have 0.082 bits higher entropy than composites; under QUARTER the direction reverses (−0.075 bits) and primes sit 26% farther from the ILR centroid. Twin primes under DOWN are 44% closer than random lag-2 pairs (distance ratio 0.56), while all other policies push them *apart* (ratios 1.22–1.38). This policy-dependent reversal confirms that the embedding faithfully reflects the interaction between number-theoretic structure and the witness selection mechanism.

12 Conjecture and open questions

We close by offering one conjecture, three questions, and several other possible directions to explore.

[Dimension continuity] The intrinsic dimension of the ILR embedding under the CENTER(f) policy is a continuous function of N and f on $(1/5, 1/3)$, with a minimum near $f = 1/4$ (the QUARTER policy).

Question 1 (Witness equidistribution). The Hardy–Littlewood heuristic provides an asymptotic *count* of Goldbach (and, by extension, Lemoine) representations of N, but says nothing about the *spatial distribution* of witnesses within a constrained window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$. The random-witness result $H = 0.49$ implies that compositional signatures drawn uniformly from the Lemoine witness pool are uncorrelated across consecutive integers—a stronger statement than mere abundance. Specifically, if the witness pool were spatially clumped or contained deserts large enough to persist across neighbouring M values, even random selection would inherit those correlations and produce $H > 1/2$.

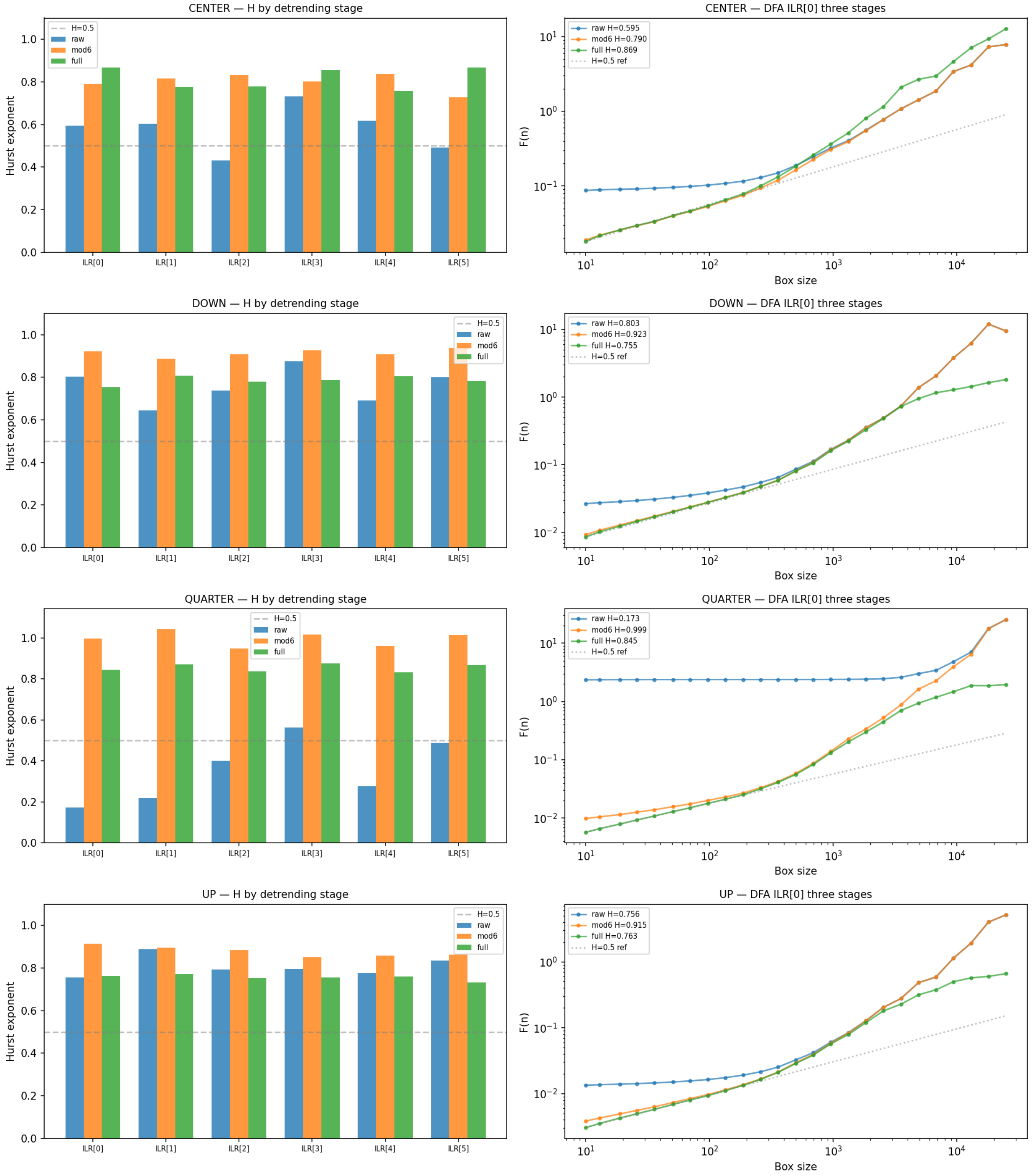


Figure 14: Three-stage DFA decomposition for deterministic policies. Left: Hurst exponents by ILR coordinate under raw (blue), mod-6 (orange), and full (green) detrending. Right: DFA log-log plots for ILR[0]. The fully-detrended slope stabilizes well above the $H = 0.5$ reference. The RANDOM policy (not shown) yields $H \approx 0.49$ at every stage and coordinate.

Does the white-noise output reflect genuine equidistribution of Lemoine pairs (p, q) within the constrained window, in the sense that their density varies smoothly with N? A positive answer would strengthen the Hardy–Littlewood counting heuristic to a spatial-regularity statement for constrained additive representations. We are not aware of existing results addressing within-window regularity for Goldbach- or Lemoine-type pairs; the extensive literature on "Goldbach numbers in short intervals" [17, 18, 19] asks whether *nearby integers* have representations, not whether the representations for a *fixed* integer are evenly spaced.

Question 2 (Policy design space). The four deterministic policies studied here are all search-start-location rules (Remark 1)—arguably the simplest possible rule set. Remaining within the constrained window, one could define policies based on rolling or periodic starting locations, modular-arithmetic rules (e.g., $q \equiv f(M) \pmod{k}$), rules conditioned on the witness chosen at the parent node, or arbitrary continuous or chaotic functions $g\colon \mathbb{R} \to [1/5, 1/3]$ composed with M. Moreover, the precise location and width of the window itself is a policy choice.

Even the simplest rules produce sharply differentiated geometry and topology. Can policies be *designed* to amplify or suppress specific arithmetic features—for instance, to maximize mod-k concentration at a target sink, or to produce embeddings with prescribed intrinsic dimension? The abundance of Lemoine and Goldbach witnesses at large scales makes the constrained window a flexible design space for probing additive and prime structure through its geometric and topological consequences.

Question 3 (Ascending construction and the achievable sublattice). The tower of admissible terminal bases (Corollary 1) can be traversed in either direction. The *ascending* construction starts from the minimal basis $\mathcal{S}_{10}$ and iteratively forms new integers as Lemoine sums $M = p + 2q$ or Goldbach sums $N = a + b$ of elements already reached, defining a growing *reachable set*

$$\begin{aligned} R_0 &= \mathcal{S}_{10}, \\ R_{k+1} &= R_k \ \cup \ \big\{\, p + 2q : p, q \in R_k \text{ prime},\ q \in [\lceil M/5 \rceil, \lfloor M/3 \rfloor],\ M := p{+}2q \,\big\} \\ &\qquad \cup \ \big\{\, a + b : a, b \in R_k \text{ odd prime} \,\big\}. \end{aligned}$$

The eventual universality of $R_\infty = \bigcup_k R_k$—that is, $R_\infty \supseteq \{N \in \mathbb{N} : N \geq 98\}$—is equivalent to the Goldbach and constrained Lemoine conjectures holding simultaneously. The ascending viewpoint does not circumvent this equivalence, but it does reframe the problem in a way that exposes new structure.

At the level of the reconstruction identity, the equation $\sum_{s \in \mathcal{S}_{10}} s \cdot W_s = N$ always has non-negative integer solutions (since $\gcd(\mathcal{S}_{10}) = 1$ and N is large). Lattice coverage is trivially universal. The gap between this and the conjecture is the *achievability constraint*: not every solution $\mathbf{W} \in \mathbb{Z}_{\geq 0}^{10}$ to the reconstruction identity is realizable as the weight vector of a descent tree, because the descent imposes a recursive factoring structure (Lemma 1) that is far more restrictive than the linear equation alone. The same gap appears at every level of the tower. The coarsening matrix C (Theorem 2) maps achievable fine-basis weight vectors to achievable coarse-basis weight vectors, but the achievable sublattice at each level is a strict subset of the reconstruction-identity solutions. A counterexample to universality—an integer N^* with no descent graph—would need to evade the achievable sublattice at every resolution level simultaneously: a tower-wide obstruction constrained by the composability of coarsening maps (Corollary 1(iii)).

Two concrete subquestions arise.

(a) Achievable sublattice characterization. For a fixed policy and basis, what is the structure of the set $\mathcal{A}_{\mathcal{S}} = \{\mathbf{W}_{\mathcal{S}}(N) : N > \max(\mathcal{S})\} \subset \mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$? This is a strict subset of the solutions to the reconstruction identity, carved out by the recursive factoring constraint. Can it be described by finitely many congruence conditions, or does it have a more complex structure?

(b) Density of the complement. Even if universal coverage cannot be proved, can the ascending construction yield a density result? The density of constrained Lemoine witnesses in the window is heuristically $\sim c\,M/\ln^2 M$. If the witness pairs for M draw from a sufficiently spread-out subset of previously reached primes—a condition closely related to the equidistribution property of Question 1—then a second-moment or sieve argument may show that $|\{N \le X : N \notin R_k\}| = o(X)$ for k growing slowly with X, i.e., that the set of integers without descent graphs has asymptotic density zero. This would be weaker than the conjecture but would follow from the tower structure (Corollary 1) combined with quantitative control on the witness distribution.

Several broader directions also merit investigation.

(Phase transitions). Related to our conjecture about dimensionality: the residue-class-dependent arcs and filaments suggest distinct basins or regimes in coefficient space. Are there sharp transitions between regimes as N or f cross certain thresholds? Sweeping the CENTER fraction over $(1/5, 1/3)$ may reveal critical boundaries at which a geometric phase transition occurs. $N/4$ may be just such a transition.

(Short memory and mixing). Quantify the rate of information decay: for each policy, how quickly does mutual information between the depth-d witness state and the terminal sink vector decay with d?

(Arithmetic fingerprints). Can the sink-coefficient embedding be used as a feature map for classification tasks, whether number-theoretic (e.g., detecting prime pairs, arithmetic progressions, or anomalous residue-class distributions) or more general applications?

(Entropy equidistribution). Does $H(\mathbf{w}(N)) \to \log_2 10$ as $N \to \infty$? That is, do the sink weights become asymptotically uniform? If so, what is the rate of convergence, and does it depend on policy?

(Mod-3 cascade proof). Can the mod-3 concentration under QUARTER be proved? The cancellation cascade (Section 8.3) has a clear heuristic basis—$p \approx 2q$ at each level produces correlated subtrees whose mod-3 residues cancel—but a rigorous statement would require control over the distribution of Lemoine witnesses near $M/4$.

(Modular recurrences). Do the residues $W_s(N) \bmod k$ satisfy recurrence relations that depend on $N \bmod m$ for structured (m, k) pairs? The reconstruction identity gives one linear constraint per modulus; the parity skeleton (Proposition 1) gives three deterministic sinks. Are there additional constraints, perhaps involving cross-sink relations, that further reduce the effective dimension of the coefficient lattice?

Acknowledgments

In preparing this manuscript, AI Large Language Models (LLMs)—ChatGPT, Google Gemini, and Claude AI—were used in the following ways: to brainstorm and scrutinize ideas; to ensure consistency and conventionality of LaTex formatting; to create code (at the author's direction) for various empirical tests and the creation of the relevant graphics; to help in identifying relevant scholarship; and to provide critical feedback and suggestions on drafts. Any AI-generated matter

has been thoroughly reviewed and edited by the author, who accepts full responsibility for the contents of this article.

A Computational details

All code is implemented in Python 3. The main computation pipeline consists of the following scripts:

- `descent_graph_sink_weights.py` for constructing Descent Graphs and computing sink-weight embeddings;
- `ilr_intrinsic_dimension.py` for ILR transformation and dimension estimation globally;
- `scale_invariance_dimension.py` for ILR transformation and dimension estimation over stratified bands;
- `persistence_homology.py` for persistent homology computation;
- `modular_arithmetic_analysis.py` for exploration of modular patterns in the raw 10-sink coefficients;
- `serial_dependence_analysis.py` for entropy stratification, autocorrelation, Hurst exponent estimation, and prime/twin-prime analysis (Section 11);
- `pseudocount_sensitivity_and_hurst_crossval.py` for pseudocount sensitivity, extended spectral analysis, and R/S cross-validation (Appendix B);
- `coarsening_matrix_verification.py` for computing the coarsening matrix $C(\mathcal{S}_{12}, \mathcal{S}_{10})$ and verifying the identity $\mathbf{W}_{\mathcal{S}_{10}}(N) = C\,\mathbf{W}_{\mathcal{S}_{12}}(N)$ (Section 10).

Primality was tested via deterministic Miller–Rabin with bases $\{2, 3, 5, 7, 11\}$, provably correct for all $n < 2.15 \times 10^{12}$. Witness lookups are memoized per (policy, M) pair for efficiency. Persistent homology uses Ripser with subsampled point clouds (800 points, 10 replicates).

Existence verification protocol. The claim that constrained witnesses exist for all integers $N \leq 10^9$ (Section 3.6) was verified by iterating over every odd integer N from 98 to 10^9 and confirming that at least one valid Lemoine witness exists in the window $[\lceil N/5 \rceil, \lfloor N/3 \rfloor]$; even integers were verified by confirming that the Goldbach scan from $\lfloor N/2 \rfloor$ succeeded. No failures were encountered. The verification script is included in the code repository.

The Python code used to generate, visualize, and analyze the Descent Graphs is available at `https://github.com/peteruphamconsulting-stack/Three-Bonsai` [8].

B Pseudocount sensitivity and structural zeros

The ILR transform requires strictly positive compositions. When some sinks receive zero mass, a pseudocount ε is added before normalization. We report both the sparsity of the raw data and the effect of varying ε across four orders of magnitude.

Sparsity. In the 100,000-integer block near 5×10^6, the fraction of structural zeros (sink weights exactly equal to zero) varies by policy: UP has no zeros at all; CENTER has 1.5% zeros, confined entirely to sink 59; DOWN has 2.3% zeros, confined entirely to sink 97; QUARTER has 23% zeros, with sink 59 identically zero for every integer (reflecting the fact that QUARTER's witness selection near $N/4$ never routes mass to the 59-basin) and sinks 2, 37, 73, and 97 each zero for one-third of integers (the N mod 3 classes that map to different islands).

Sensitivity sweep. We sweep $\varepsilon \in \{10^{-3}, 10^{-2}, 0.1, 0.5, 1.0, 2.0, 5.0\}$ and recompute three key quantities: kNN-MLE intrinsic dimension, PCA variance concentration (fraction explained by the first three components), and the DFA(1) Hurst exponent of the first ILR coordinate (mod-6 detrended).

Table 5: Sensitivity of key geometric quantities to the pseudocount ε. Dimension range is over the seven ε values tested. All data from 100,000 consecutive integers near 5×10^6.

Policy	Dim. range	PCA-3 range	H_{DFA} range
CENTER	4.14–4.24	0.9844–0.9953	0.783 (invariant)
DOWN	4.50–4.68	0.9978–0.9994	0.904 (invariant)
QUARTER	3.00–3.10	0.9994–0.9999	1.001 (invariant)
UP	5.04–5.20	0.8889–0.8891	0.911 (invariant)

Across 3.5 orders of magnitude of ε, the intrinsic dimension fluctuates by at most 0.20 (well within the estimation variance of kNN-MLE), the PCA concentration is essentially unchanged, and the Hurst exponent is *completely invariant* to three decimal places (Table 5; Figure 15).

Why the pseudocount does not matter. The sink weights $\mathbf{W}(N)$ are *lattice compositional data*—integer counts subject to the constant-sum constraint (3)—rather than the continuous proportions for which the ILR transform was originally designed. In the statistical literature on count compositions, applying log-ratio transforms to discrete data is known to introduce additive quantization variation that can distort distances when counts are low. In our setting this distortion is negligible for a structural reason: the embedding produces coefficients typically in the hundreds or thousands. The sparse sinks (e.g., $W_{59} = 0$ under QUARTER) are precisely those whose contribution to the Helmert orthonormal projection is most suppressed by the mass distribution. The macroscopic topology of the ILR manifold is therefore invariant to additive pseudocount perturbation at all sinks.

Despite QUARTER's 23% sparsity rate, the pseudocount has no measurable effect on its geometry or memory structure. This validates the use of the default $\varepsilon = 0.5$ throughout the paper.

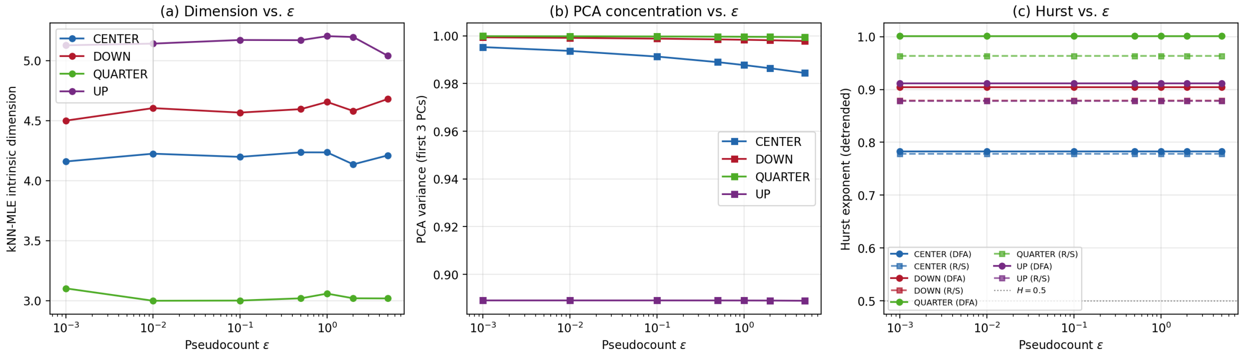


Figure 15: Sensitivity of intrinsic dimension, PCA variance concentration, and Hurst exponent to the pseudocount ε. All three quantities are stable across four orders of magnitude.

C Serial dependence analysis details

The serial dependence experiments (Section 11) use a block of 100,000 consecutive integers ($N =$ 5,000,001 to 5,100,000). Hurst exponents are estimated via DFA(1) with box sizes logarithmically spaced from 10 to $T/4$. The shuffle null permutes ILR vectors within each mod-6 residue class (50 shuffles).

References

[1] J. Aitchison, *The Statistical Analysis of Compositional Data*, Chapman and Hall, London, 1986.

[2] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, American Mathematical Society, 2010.

[3] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D* **9** (1983), 189–208.

[4] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," *Advances in Neural Information Processing Systems* **17** (2004).

[5] T. Oliveira e Silva, S. Herzog, and S. Pardi, "Empirical verification of the even Goldbach conjecture and computation of prime gaps up to 4×10^{18}," *Mathematics of Computation* **83** (2014), 2033–2060.

[6] C. Tralie, N. Saul, and R. Bar-On, "Ripser.py: A lean persistent homology library for Python," *Journal of Open Source Software* **3**(29) (2018), 925.

[7] P. Upham, *Deterministic Smooth Number Cover and Lemoine Obstructions*, doi:10.5281/zenodo.18226015 (2025).

[8] P. Upham, *Three-Bonsai: Code for Goldbach–Lemoine Descent Graph analysis*, GitHub repository, `https://github.com/peteruphamconsulting-stack/Three-Bonsai` (2026).

[9] H. Levy, "On Goldbach's conjecture," *Math. Gazette* **47** (1963), 274.

[10] Z. Juhász, M. Bartalos, P. Magyar, and G. Farkas, "Empirical verification of a generalization of Goldbach's conjecture," *Journal of Integer Sequences* **27**(4) (2024), Article 24.4.3.

[11] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society* **46**(2) (2009), 255–308.

[12] H. E. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers* **116** (1951), 770–808.

[13] S. Das, P. Ghosh, S. Ghosh, and S. Sen, "Oriented bipartite graphs and the Goldbach graph," *Discrete Mathematics* **344**(10) (2021), Article 112497; arXiv:1611.10259.

[14] Y. Fujiyoshi, "On the finiteness of prime trees and their relation to modular forms," arXiv:2601.23016 (2026).

[15] M. E. Houle, E. Schubert, and A. Zimek, "On the correlation between local intrinsic dimensionality and outlierness," *Similarity Search and Applications (SISAP)*, Springer LNCS 11223 (2018), 177–191.

[16] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E* **49**(2) (1994), 1685–1689.

[17] H. L. Montgomery and R. C. Vaughan, "The exceptional set in Goldbach's problem," *Acta Arithmetica* **27** (1975), 353–370.

[18] A. Languasco, "A singular series average and Goldbach numbers in short intervals," *Acta Arithmetica* **83** (1998), 171–179.

[19] A. Chirre and M. V. Hagen, "On Goldbach numbers in short intervals," arXiv:2512.23534 (2025).

[20] L. Wang, Y. Huang, and Z.-C. Dai, "Fractal in the statistics of Goldbach partition," arXiv:nlin/0601024 (2006).

[21] O. Shanker, "Random matrices, generalised zeta functions and self-similarity of zero distributions," *Journal of Physics A: Mathematical and General* **39**(45) (2006), 13983–13997.