Neruthes Necklace: A Balanced Binary Recursion and Its Binary-Tree Automorphisms

Neruthes
2026-11-07

Abstract

We introduce the **Neruthes necklace**, a recursively constructed sequence of dyadic subsets of the unit interval. Its defining operation acts on adjacent pairs of cells containing exactly one occupied cell each, and replaces each such pair by a four-bit word containing exactly two occupied cells. Starting from the initial configuration 10, the replacement is selected by a deterministic binary sequence, called the Coin sequence, and is applied separately to each parent pair from left to right. Thus the necklace is defined by a simple local rule, while its successive stages form a non-nested sequence of sets.

The construction has an exact finite-stage balance: if A_n denotes the occupied set at level n, then $\lambda(A_n) = \frac{1}{2}$ for every n. Starting from this elementary geometric observation, we investigate the structure that emerges across scales. We prove that, on the Lebesgue probability space $([0,1), \mathcal{B}, \lambda)$, the membership indicators $\mathbb{1}_{A_n}$ form an independent sequence of fair Bernoulli random variables. Consequently, $\lambda(\limsup_n A_n) = 1$, $\lambda(\liminf_n A_n) = 0$, and the Cesàro averages converge almost everywhere to $\frac{1}{2}$.

The point-set structure requires a distinction between symbolic paths and their interval representations. We prove that the symbolic eventual-membership sets are countably infinite and dense in the binary sequence space. Their projections to the interval are also countably infinite and dense, with Hausdorff and packing dimensions zero and both box dimensions one. The actual interval liminf, however, is obtained only after restricting to canonical binary expansions. It is at most countable, but its density and cardinality depend on the Coin sequence. In particular, for the constant-one Coin sequence, the interval liminf is empty.

We then identify the binary-tree structure underlying the recursion. The membership indicators can be represented as the output coordinates of a triangular, prefix-dependent bit transformation. This transformation is a rooted binary tree automorphism and preserves the fair-coin product measure. More strongly, any two Coin sequences give rise to necklaces that are exactly conjugate, level by level, by a measure-preserving isometry of the standard binary ultrametric space. Thus the Coin sequence changes the interval realization of the necklace, but not its intrinsic binary-tree metric structure.

We further construct an explicit family of finite-state Coin sequences indexed by integers $k \geq 2$. For this family, the interval liminf is a countable set whose closure is a compact symbolic fractal defined by the avoidance of the word 0^k. We prove that the limsup is the entire interval $[0,1)$, while the liminf has Hausdorff and packing dimensions zero and both box dimensions equal to $d_k = \log_2(\lambda_k)$, where λ_k is the positive root of the

corresponding k-step recurrence. The compact closure has Hausdorff dimension d_k as well. These dimensions form a strictly increasing sequence converging to one.

Beyond the universal tree structure, the finite-stage Euclidean geometry depends on the Coin sequence. As a first quantitative invariant, we derive an exact formula for the number R_n of connected components of A_n in terms of the parent-pair orientations. We then compute R_n exactly for two families: the alternating Coin sequence, where $R_n = 2^{n-2}$ for $n \geq 2$, and the constant-one Coin sequence, where $R_n = \frac{2^n+(-1)^{n-1}}{3}$. The latter formula follows from an explicit identification of the parent-pair orientation words with the Thue–Morse sequence and its complement.

The purpose of this article is to develop the Neruthes necklace from its local recursive definition and to derive its probabilistic, measure-theoretic, and tree-theoretic properties. We also formulate further questions about the extent to which different Coin sequences can produce distinct geometries when the binary ultrametric is replaced by the usual Euclidean metric on the interval.

Contents

Abstract 1
1 Introduction 3
2 The Construction 5
2.1 Dyadic cells and configurations 5
2.2 Balanced pairs 6
2.3 The Expand operation 6
2.4 Variants of the Expansion Mechanism 7
2.5 Basic invariants 7
3 The Probabilistic Skeleton 8
3.1 The triangular bit representation 8
3.2 Independence 8
3.3 Weak convergence 9
4 The Almost-Sure and Pointwise Structure 10
4.1 Almost-sure frequency 10
4.2 Countability of the interval exceptional sets 11
4.3 Symbolic eventual-membership sets 11
4.4 Dimension of the projected symbolic set 12
4.5 An endpoint-sensitive example 13
4.6 Self-Encounters with the Coin Sequence 14
4.7 Summary 14
5 Binary Tree Structure and Equivalence 15
5.1 The binary sequence space 15
5.2 The tree automorphism 15
5.3 Conjugacy between different Coin sequences 16
5.4 An Explicit Interval Example 17
5.5 What Changes under Interval Coding? 18
5.6 Equivalence questions 19
5.7 The classification problem 20
6 Finite-Stage Geometry and Special Coin Sequences 20

6.1 An Exact Component-Count Formula 20
6.2 The Alternating Coin Sequence 22
6.3 An Exact Formula for the Constant-One Coin Sequence 23
6.4 The Thue–Morse Coin Sequence 26
6.5 A Finite-Level Comparison 26
6.6 The Induced Thue–Morse Orientation Words 27
6.7 Component-Count Formulas and the Modulo-Four Pattern 28
6.8 Scope of the Geometric Statistic 31
7 Visualization 31
7.1 Constant Coin Sequences 31
7.1.1 All True 31
7.1.2 All False 31
7.2 Periodic Coin Sequences 31
7.2.1 Alternating 10 31
7.2.2 Alternating 01 31
7.3 A Non-periodic Coin Sequence 31
7.3.1 Thue–Morse 31
8 Conclusion and Open Problems 32
8.1 Established results 32
8.2 Open Problems 33
8.3 A Longer-Term Interface 36
8.4 Final perspective 36
9 Acknowledgment 36

1 Introduction

Many classical fractal constructions are based on nested approximations. The Cantor set, for example, is obtained from a decreasing sequence of compact sets, and each successive stage retains the geometric information of the preceding one. Such constructions naturally lead to limit sets defined by intersections, and their geometry is closely tied to the refinement process.

The construction studied in this article begins with a different principle. We introduce the **Neruthes necklace**, a recursively generated sequence of dyadic subsets of the unit interval. At every finite stage, exactly half of the dyadic cells are occupied. The next stage is obtained by replacing each adjacent pair of cells with a balanced four-cell pattern. The replacement is selected by a deterministic binary sequence, called the **Coin sequence**. The rule is local, but its repeated application creates a non-nested sequence of sets.

The elementary balance of the construction conceals a richer structure. Although every stage has the same Lebesgue measure, the occupied regions move across scales. This raises several natural questions. What can be said about the points that belong to infinitely many stages? Which points eventually belong to every stage? Does the deterministic choice of replacements affect the resulting limit sets? And what mathematical structure explains the probabilistic behaviour that appears when the construction is viewed on the Lebesgue probability space?

Our first objective is to answer these questions directly from the recursive definition of the Neruthes necklace. We show that the membership indicators of its stages are independent fair Bernoulli random variables, for every deterministic Coin sequence. This gives a complete description of several measure-theoretic properties, including the measures of the limsup and liminf sets and the almost-everywhere limiting frequency of membership.

The point-set structure is more delicate. We distinguish the symbolic eventual-membership sets from their interval realizations. In the binary sequence space, the eventual-one and eventual-zero sets are countably infinite and dense. Their images under the binary coding map are also countably infinite and dense in $[0,1]$, with Hausdorff and packing dimensions zero and both lower and upper box dimensions one.

The actual interval liminf, however, is obtained only from canonical binary expansions. It is at most countable, but its density and cardinality are not universal. In particular, for the constant-one Coin sequence, the interval liminf is empty. Thus symbolic density does not automatically transfer to the interval liminf.

A further structural feature emerges when the recursion is encoded by binary expansions. The membership indicators can be expressed as a triangular transformation of the binary digits of a point. This transformation is a rooted binary tree automorphism. In particular, different Coin sequences do not produce different intrinsic geometries in the standard binary ultrametric: their necklaces are related by an explicit level-preserving isometry. This observation refines the role of the Coin sequence. It changes the concrete realization of the necklace in the interval, but it does not change its underlying binary-tree metric structure.

A further result shows that the balanced recursion can generate a family of fractal closures with exact, parameter-dependent dimensions. We construct finite-state Coin sequences satisfying the complementary-sibling condition required by the original Expand rule. The local geometric replacement rule remains unchanged throughout the family; only a finite-state admissibility constraint, parametrized by an integer $k \geq 2$, is varied. The admissible-prefix growth rate is governed by a Perron root λ_k, and the resulting compact symbolic closure has Hausdorff and box dimensions

$$d_k = \log_2(\lambda_k),$$

where

$$\lambda_k^k = \lambda_k^{k-1} + \ldots + \lambda_k + 1.$$

The dimensions d_k are strictly increasing and converge to one. Meanwhile, the corresponding interval liminf sets are countable, have Hausdorff and packing dimensions zero, and have both lower and upper box dimensions equal to d_k. Thus the construction produces exact dimension formulae from finite-state symbolic constraints within the original balanced recursion, rather than prescribing the geometric contraction ratios independently.

The article also investigates the finite-stage Euclidean geometry of the necklace. We use the number of connected components as a finite-stage invariant and derive an exact formula in terms of the orientation word generated by the recursion. This formula reduces component counting to a transition statistic of a binary word. We then obtain exact results

for the alternating and constant-one Coin sequences and examine the corresponding structure for the Thue–Morse Coin sequence. These computations motivate a modulo-four coincidence conjecture relating the component counts of the three special sequences. The conjecture is supported by finite-level computations, but is presented as an open problem rather than an established theorem.

The article is organized as follows. Section 2 gives the precise definition of the Neruthes necklace and establishes its finite-stage balance. Section 3 identifies the triangular bit transformation and proves the independence theorem. Section 4 derives the measure-theoretic consequences and distinguishes symbolic eventual-membership sets from their interval realizations. Section 5 develops the binary-tree interpretation and proves the exact conjugacy between different Coin sequences.

Section 6 constructs a family of finite-state Coin sequences and characterizes the associated interval liminf sets. It also computes the Hausdorff and box dimensions of these exceptional sets and their compact closures. Section 7 studies finite-stage component counts and special Coin sequences. It establishes the general component-count formula, derives exact formulae for the alternating and constant-one sequences, and investigates the Thue–Morse orientation structure and the resulting modulo-four conjecture. Section 8 discusses connections with substitution dynamics, balanced-pair algorithms, and related geometric constructions. Section 9 presents visualizations of the construction for constant, periodic, and non-periodic Coin sequences. Finally, Section 10 collects the established results and open problems.

The probabilistic consequences below are classical once independence has been established. Our purpose is not to claim a new probability law, but to explain how this law is encoded by the Neruthes necklace, how it coexists with a distinctive point-set structure, and how both the exceptional-set dimensions and the finite-stage Euclidean geometry reflect the deterministic Coin sequence.

2 The Construction

2.1 Dyadic cells and configurations

For each integer $n \geq 1$ and each index $i \in \{0, 1, ..., 2^n - 1\}$, define the dyadic cell

$$I_{n,i} = \left[\frac{i}{2^n}, \frac{i+1}{2^n}\right).$$

A level-n configuration is a binary word

$$b^{(n)} = \left(b_0^{(n)}, b_1^{(n)}, ..., b_{2^n-1}^{(n)}\right) \in \{0,1\}^{2^n}.$$

The corresponding occupied set is

$$A_n = \cup_{i:b_i^{(n)}=1} I_{n,i}.$$

We start from the initial configuration

$$b^{(1)} = (1, 0),$$

so that

$$A_1 = \left[0, \frac{1}{2}\right).$$

At every later stage, the configuration is obtained by applying the Expand operation to all adjacent pairs of the preceding configuration.

2.2 Balanced pairs

At level n, group the entries of $b^{(n)}$ into adjacent pairs:

$$p_j = \left(b_{2j}^{(n)}, b_{2j+1}^{(n)}\right), \quad 0 \le j < 2^{n-1}.$$

The corresponding parent interval is

$$P_{n,j} = I_{n,2j} \cup I_{n,2j+1} = \left[\frac{j}{2^{n-1}}, \frac{j+1}{2^{n-1}}\right).$$

The initial configuration consists of the balanced pair 10. Moreover, both possible output words of the Expand operation, 0110 and 1001, split into the balanced pairs $01, 10$ and $10, 01$, respectively. It follows by induction that every adjacent pair at every level is either 01 or 10.

In particular, each parent pair contains exactly one occupied cell.

2.3 The Expand operation

Let

$$\sigma = (\sigma_1, \sigma_2, ...) \in \{0,1\}^{\mathbb{N}}$$

be a fixed deterministic binary sequence, called the **Coin sequence**. At level n, the operation processes the parent pairs from left to right. The pair with index j uses the Coin value

$$\sigma_{2^{n-1}+j}, \quad 0 \le j < 2^{n-1}.$$

The replacement rule is

$$01 \to \begin{cases} 1001 \\ \qquad \text{if} \;\; \sigma_{2^{n-1}+j} = 1 \\ 0110 \\ \qquad \text{if} \;\; \sigma_{2^{n-1}+j} = 0 \end{cases}$$

and

$$10 \to \begin{cases} 0110 \\ \qquad \text{if} \;\; \sigma_{2^{n-1}+j} = 1 \\ 1001 \\ \qquad \text{if} \;\; \sigma_{2^{n-1}+j} = 0. \end{cases}$$

Thus every parent pair is replaced by one of the two words 0110 and 1001. Each output word contains exactly two ones and two zeros.

Write $w_j = \big(w_{j,0}, w_{j,1}, w_{j,2}, w_{j,3}\big)$ for the word produced from the j-th parent pair. The next configuration is defined by

$$b_{4j+s}^{(n+1)} = w_{j,s}, \quad s \in \{0,1,2,3\}.$$

The four entries occupy, in their original order, the four consecutive dyadic grandchildren of $P_{n,j}$. No global rearrangement of cells takes place.

2.4 Variants of the Expansion Mechanism

The expansion mechanism described above is one particular realization of a more general local substitution scheme. We call the local mapping from a parent pair and a Coin bit to a four-bit output block an **expansion kernel**. Thus, an expansion kernel is a map

$$K : \{01,10\} \times \{0,1\} \to \{0101, 0110, 1001, 1010\}.$$

The kernel specified in this paper is not the only possible choice. In particular, there are eight bijective kernels compatible with the rooted binary-tree hierarchy, each assigning the four possible input states to the four admissible output blocks. Every such output block contains exactly two zeros and two ones, so the blockwise balance property is preserved.

The Coin-reading convention is another independent design choice. The starting position used to read the Coin sequence at each level may be prescribed by a user-defined function $\mathrm{GetStart}(n)$. The reading policies adopted in this paper are specific choices within this broader framework. Other policies can be implemented without changing the general architecture of the expansion procedure.

The results established in this paper concern the particular kernel and Coin-reading convention explicitly specified above. Whether the probabilistic, symbolic, and geometric conclusions extend to other choices is a separate question.

2.5 Basic invariants

Every parent pair produces exactly two occupied children. Since there are 2^{n-1} parent pairs at level n, the number of occupied cells at level n is

$$\mathrm{card}\Big(\Big\{i : b_i^{(n)} = 1\Big\}\Big) = 2^{n-1}.$$

for every $n \geq 1$, and hence

$$\lambda(A_n) = 2^{n-1} \cdot 2^{-n} = \frac{1}{2}.$$

The same induction also shows that every adjacent pair in every configuration is balanced. Equivalently,

$$b_{2j}^{(n)} + b_{2j+1}^{(n)} = 1$$

for all admissible n and j.

The sequence $(A_n)_{n\geq 1}$ is generally neither increasing nor decreasing. In particular, the occupied grandchildren need not lie inside the occupied child of their parent pair. The construction therefore preserves the total occupied measure without preserving inclusion between successive stages.

This local balance is the defining feature of the Neruthes necklace. The measure-theoretic and tree-theoretic properties developed in the following sections will be derived from this recursive rule.

3 The Probabilistic Skeleton

For $x \in [0,1)$, choose its binary expansion

$$x = 0.x_1x_2x_3\ldots_2,$$

using the convention that dyadic rationals have expansions ending in zeros. The level-n cell containing x is determined by the prefix $(x_1, \ldots, x_n)$, and

$$\mathbb{1}_{A_n}(x) = b^{(n)}_{\lfloor 2^n x \rfloor}.$$

The balanced-pair property gives a simple description of this membership indicator.

3.1 The triangular bit representation

Lemma 1. For every $n \geq 1$, there exists a function

$$\varphi_n : \{0,1\}^{n-1} \to \{0,1\}$$

such that, for every $x \in [0,1)$,

$$\mathbb{1}_{A_n}(x) = x_n \oplus \varphi_{n(x_1,\ldots,x_{n-1})}.$$

Here φ_1 is a function on the singleton set containing the empty word.

Proof. Fix a prefix $p = (x_1, \ldots, x_{n-1})$. The two level-$n$ cells extending this prefix correspond to the two possible values of the final binary digit x_n. They form one adjacent pair in the level-n configuration.

By the balanced-pair property, exactly one of these two cells is occupied. Therefore, as a function of the final digit x_n, the membership indicator takes the values 0 and 1 in some order. There is consequently a unique value $\varphi_{n(p)} \in \{0,1\}$ for which the indicator is equal to $x_n \oplus \varphi_{n(p)}$.

Since the canonical binary-expansion convention is consistent with the half-open dyadic cells, the identity holds also at dyadic endpoints. □

3.2 Independence

Theorem 1 (Independence). For every deterministic Coin sequence σ, the sequence

$$\left(\mathbb{1}_{A_n}\right)_{n\geq 1}$$

is a sequence of independent fair Bernoulli random variables on the Lebesgue probability space $([0,1), \mathcal{B}, \lambda)$.

Proof. Let

$$Y_{n(x)} = \mathbb{1}_{A_n}(x).$$

By Lemma 1,

$$Y_n = x_n \oplus \varphi_{n(x_1,\ldots,x_{n-1})}.$$

Under Lebesgue measure, the binary digits $x_1, x_2, \ldots$ are independent fair Bernoulli variables, apart from the irrelevant ambiguity at dyadic rationals.

For each n, the value of $\varphi_{n(x_1,\ldots,x_{n-1})}$ is determined by the first $n-1$ binary digits. Conditional on these digits, x_n is still a fair Bernoulli variable. Flipping a fair bit by a fixed value does not change its distribution. Thus

$$\lambda(Y_n = 1 \mid x_1, \ldots, x_{n-1}) = \frac{1}{2}$$

and

$$\lambda(Y_n = 0 \mid x_1, \ldots, x_{n-1}) = \frac{1}{2}.$$

In particular, Y_n is independent of the sigma-algebra generated by the first $n-1$ binary digits. Since $Y_1, \ldots, Y_{n-1}$ are measurable with respect to that sigma-algebra, Y_n is independent of the preceding membership indicators.

Each Y_n is fair, and each is independent of all its predecessors. Hence the sequence is independent and identically distributed, with

$$\lambda(Y_n = 0) = \lambda(Y_n = 1) = \frac{1}{2}.$$

This proves the theorem. □

The theorem identifies the probabilistic structure hidden in the Neruthes necklace recursion. The Coin sequence determines the prefix-dependent flips φ_n, but it does not affect the joint distribution of the membership indicators. In this sense, the deterministic recursion produces the same Bernoulli process for every choice of Coin sequence.

3.3 Weak convergence

Let $f_n = \mathbb{1}_{A_n}$ and $g_n = f_n - \frac{1}{2}$.

By Theorem 1, the functions g_n are pairwise orthogonal in $L^2([0,1])$, and

$$\|g_n\|_2 = \frac{1}{2}.$$

For any $h \in L^2([0,1])$, Bessel's inequality therefore implies

$$\int_0^1 h(x) g_{n(x)} \, \mathrm{dx} \to 0.$$

Thus g_n converges weakly to zero in L^2.

Since the functions g_n are uniformly bounded, the convergence against bounded test functions extends by a truncation argument to all test functions in $L^{q([0,1])}$, where $1 < q < \infty$.

Consequently, f_n converges weakly to $\frac{1}{2}$ in $L^{p([0,1])}$ for every $1 < p < \infty$.

Similarly, bounded functions are dense in $L^1([0,1])$. The same argument therefore gives weak-* convergence in $L^{\infty([0,1])}$.

These convergences are not strong convergences. Indeed, since f_n is an indicator function and $\lambda(A_n) = \frac{1}{2}$, we have

$$\left\| f_n - \frac{1}{2} \right\|_p = \frac{1}{2} \quad \text{for every} \quad 1 \leq p \leq \infty.$$

Thus f_n does not converge strongly to $\frac{1}{2}$ in any of these L^p spaces.

4 The Almost-Sure and Pointwise Structure

Throughout this section, write

$$Y_{n(x)} = \mathbb{1}_{A_n}(x) = x_n \oplus \varphi_{n(x_1,\ldots,x_{n-1})}.$$

We distinguish the measure-theoretic behaviour of the membership indicators from the pointwise structure of their eventual behaviour.

4.1 Almost-sure frequency

Proposition 1 (Almost-sure frequency).

For every deterministic Coin sequence,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{A_n}(x) = \frac{1}{2}$$

for Lebesgue-almost every $x \in [0,1)$.

Consequently,

$$\lambda\Big(\limsup_n A_n\Big) = 1$$

and

$$\lambda\Big(\liminf_n A_n\Big) = 0.$$

Proof.

By Theorem 1, the membership indicators are independent Bernoulli random variables with common mean $\frac{1}{2}$.

The strong law of large numbers therefore gives the asserted almost-everywhere convergence.

If x belongs to only finitely many A_n, the corresponding averages converge to zero. If x belongs to every sufficiently late A_n, the averages converge to one.

Both possibilities contradict the almost-everywhere limit $\frac{1}{2}$. Hence the complement of the limsup and the liminf are null sets, respectively.

This proves the proposition. □

4.2 Countability of the interval exceptional sets

Theorem 2 (Countability).

For every deterministic Coin sequence, both sets

$$\liminf_n A_n$$

and

$$[0,1) \setminus \limsup_n A_n$$

are at most countable.

Proof.

Fix $N \geq 1$.

Suppose that $Y_{n(x)} = 1$ for every $n \geq N$. Once the first $N-1$ binary digits have been chosen, the equation

$$x_n = 1 \oplus \varphi_{n(x_1,\ldots,x_{n-1})}$$

determines every subsequent digit uniquely.

There are only 2^{N-1} possible prefixes of length $N-1$. Thus the set of points satisfying this eventual condition is finite.

Taking the union over all N proves that $\liminf_n A_n$ is at most countable.

The same argument, using

$$x_n = \varphi_{n(x_1,\ldots,x_{n-1})},$$

shows that the set of points for which $Y_{n(x)} = 0$ eventually is at most countable. This set is exactly $[0,1) \setminus \limsup_n A_n$.

The theorem follows. □

Corollary 1.

The set of points at which the sequence $\left(\mathbb{1}_{A_n}(x)\right)_n$ converges is at most countable.

Proof.

A binary sequence converges if and only if it is eventually constant. The convergence set is therefore the union of the two countable sets considered in Theorem 2. □

4.3 Symbolic eventual-membership sets

The preceding countability statement concerns the interval with its specified binary-expansion convention. We now consider the full binary sequence space

$$\Omega = \{0,1\}^{\mathbb{N}}.$$

Define the triangular map

$$T_{\sigma(\beta)_n} = \beta_n \oplus \varphi_n^{\sigma(\beta_1,\ldots,\beta_{n-1})}.$$

For each fixed prefix, the next output digit is obtained either by preserving or by flipping the next input digit. The map is consequently a bijection on every finite level of the binary tree, and hence a homeomorphism of Ω.

Let

$$\mathcal{E}_1^\sigma = \Big\{\beta \in \Omega : T_{\sigma(\beta)_n} = 1 \text{ for all sufficiently large } n\Big\},$$

and define $\mathcal{E}_0^\sigma$ analogously.

Theorem 3 (Symbolic eventual sets).

For every Coin sequence, both $\mathcal{E}_1^\sigma$ and $\mathcal{E}_0^\sigma$ are countably infinite and dense in Ω.

Proof.

The set of binary sequences that are eventually equal to one is countably infinite and dense in Ω. The same is true of the set of sequences eventually equal to zero.

Since T_σ is a homeomorphism, the inverse images of these two sets have the same cardinality and density.

These inverse images are precisely $\mathcal{E}_1^\sigma$ and $\mathcal{E}_0^\sigma$. $\square$

Let

$$\pi(\beta) = \sum_{n=1}^{\infty} \beta_n 2^{-n}$$

be the binary coding map, and let

$$\Omega_{\text{can}} = \{\beta \in \Omega : \beta \text{ is not eventually equal to one}\}.$$

This is the set of canonical expansions of points in $[0, 1)$.

The actual interval liminf satisfies

$$\liminf_n A_n = \pi(\mathcal{E}_1^\sigma \cap \Omega_{\text{can}}).$$

Similarly,

$$[0, 1) \setminus \limsup_n A_n = \pi(\mathcal{E}_0^\sigma \cap \Omega_{\text{can}}).$$

In particular, the density of the symbolic eventual sets does not imply that their intersections with Ω_{can} are dense.

4.4 Dimension of the projected symbolic set

Define

$$D_1^\sigma = \pi(\mathcal{E}_1^\sigma).$$

Theorem 4 (Dimension of the projected symbolic sets).

For every Coin sequence, both D_1^σ and D_0^σ are countably infinite and dense in $[0, 1]$. Moreover, for $i \in \{0, 1\}$,

$$\dim_{H(D_i^\sigma)} = \dim_{P(D_i^\sigma)} = 0,$$

whereas

$$\underline{\dim}_{B(D_i^\sigma)} = \overline{\dim}_{B(D_i^\sigma)} = 1.$$

Proof.

By Theorem 3, $\mathcal{E}_1^\sigma$ is countably infinite and dense in Ω.

The binary coding map is continuous and surjective, with fibers containing at most two sequences. Its image D_1^σ is therefore countably infinite.

Continuity and surjectivity imply that the image of a dense subset is dense in $[0,1]$.

Countability gives Hausdorff and packing dimensions equal to zero. Density gives

$$N(D_1^\sigma, r) = \Theta\left(\frac{1}{r}\right),$$

where $N(E,r)$ is the minimum number of intervals of diameter at most r needed to cover E.

Hence both box dimensions are one. □

The set D_1^σ is a projection of the symbolic eventual-membership locus. It should not be identified with the interval liminf, which is obtained only after restricting to canonical binary expansions.

4.5 An endpoint-sensitive example

For the constant-one Coin sequence $\sigma = (1,1,1,...)$, the prefix-dependent flip is

$$\varphi_n^{\sigma(\beta)} = P_{n-1}(\beta) \oplus (n \bmod 2),$$

where

$$P_{n-1}(\beta) = \left(\sum_{r=1}^{n-1} \beta_r\right) \bmod 2.$$

Suppose that $Y_{n(\beta)}$ is eventually equal to a constant $c \in \{0,1\}$. Then

$$\beta_n = c \oplus P_{n-1}(\beta) \oplus (n \bmod 2).$$

It follows that

$$P_{n(\beta)} = c \oplus (n \bmod 2).$$

Consequently, for every sufficiently large n,

$$\beta_n = c \oplus c \oplus ((n-1) \bmod 2) \oplus (n \bmod 2) = 1.$$

Thus every path whose membership indicators are eventually constant is itself eventually equal to one.

Such a path is not a canonical binary expansion of a point in $[0,1)$. Therefore, for the constant-one Coin sequence,

$$\liminf_n A_n = \varnothing$$

and

$$\limsup_n A_n = [0, 1).$$

This example shows why symbolic eventual behaviour and interval pointwise behaviour must be distinguished.

4.6 Self-Encounters with the Coin Sequence

We briefly consider whether the output of the expansion can reproduce an initial segment of the Coin sequence from which it was generated.

For each level n, let L_n be the binary word obtained by decoding the configuration $b^{(n)}$ in consecutive pairs, according to $01 \mapsto 0$ and $10 \mapsto 1$. We say that a self-encounter occurs at level n if

$$L_n = (\sigma_1, ..., \sigma_{2^{n-1}}).$$

The defining blocks 0110 and 1001 decode respectively to 01 and 10. Consequently, every consecutive pair of bits in L_n consists of complementary bits. Thus, a self-encounter at level n requires

$$(\sigma_{2j-1}, \sigma_{2j}) \in \{(0, 1), (1, 0)\}, \qquad 1 \leq j \leq 2^{n-2}.$$

Under the fair-coin product measure, these 2^{n-2} disjoint pairs are independent, and each is complementary with probability $1/2$. Hence

$$\mathbb{P}(\text{self-encounter at level } \ n) \leq 2^{-2^{n-2}}.$$

Since

$$\sum_{n=2}^{\infty} 2^{-2^{n-2}} < \infty,$$

the first Borel–Cantelli lemma implies that, for almost every Coin sequence, self-encounters occur at only finitely many levels.

This gives an almost-sure finiteness property for self-encounters, without requiring independence between the self-encounter events at different levels.

4.7 Summary

For every deterministic Coin sequence, the membership indicators are independent fair Bernoulli variables. Their empirical averages converge almost everywhere to $\frac{1}{2}$, and

$$\lambda\Big(\limsup_n A_n\Big) = 1, \quad \lambda\Big(\liminf_n A_n\Big) = 0.$$

The interval liminf and the complement of the interval limsup are at most countable, but their density and exact cardinality depend on the Coin sequence. Thus density and box dimension one are properties of the symbolic or projected sets, not of the interval liminf in general.

In contrast, the corresponding symbolic eventual-membership sets are always countably infinite and dense in the binary sequence space.

5 Binary Tree Structure and Equivalence

The preceding sections show that the measure-theoretic properties of the Neruthes necklace do not depend on the Coin sequence. We now identify the underlying structure responsible for this invariance.

The natural setting is the space of infinite binary sequences. In this space, the recursive construction becomes a triangular transformation of binary digits, and the transformations associated with different Coin sequences can be compared exactly.

5.1 The binary sequence space

Let

$$\Omega = \{0,1\}^{\mathbb{N}}$$

be the space of infinite binary sequences. Equip Ω with the fair-coin product measure μ and the standard binary ultrametric

$$d(x,y) = 2^{-k},$$

where k is the first coordinate at which x and y differ. We set $d(x,x) = 0$ for every $x \in \Omega$.

A cylinder determined by a finite word $p = (p_1, ..., p_m)$ is

$$[p] = \{x \in \Omega : x_1 = p_1, ..., x_m = p_m\}.$$

These cylinders form a basis for the topology of Ω, and each has measure 2^{-m}.

For a fixed Coin sequence σ, define

$$T_{\sigma(x)_n} = x_n \oplus \varphi_n^{\sigma(x_1,\dots,x_{n-1})},$$

where φ_n^{σ} is the prefix-dependent bit-flip function associated with the Neruthes necklace recursion.

5.2 The tree automorphism

Theorem 5 (Tree automorphism). For every Coin sequence σ, the map

$$T_\sigma : \Omega \to \Omega$$

is a rooted binary tree automorphism. In particular, it is a homeomorphism, an isometry for the binary ultrametric, and a measure-preserving transformation of (Ω, μ).

Proof. The first n output digits of $T_{\sigma(x)}$ depend only on the first n input digits. Thus T_σ maps each cylinder of length n into a cylinder of length n.

More explicitly, suppose the first $n-1$ input digits have already been specified. Then the last output digit is

$$y_n = x_n \oplus \varphi_n^{\sigma(x_1,\dots,x_{n-1})}.$$

For the fixed prefix, this map either leaves the two possible values of x_n unchanged or interchanges them. It is therefore a bijection on the two children of every vertex of the binary tree.

By induction on the level, T_σ induces a bijection on every finite level of the rooted binary tree, preserving the parent-child relation. It is consequently a rooted binary tree automorphism.

A rooted tree automorphism preserves the length of the common prefix of any two infinite paths. Hence

$$d\big(T_{\sigma(x)}, T_{\sigma(y)}\big) = d(x, y),$$

so T_σ is an isometry. It also maps every cylinder of length n onto a cylinder of the same length. Since all such cylinders have measure 2^{-n}, the product measure is preserved. □

The construction of the Neruthes necklace is encoded by the coordinate identity

$$Y_{n(x)} = \big(T_{\sigma(x)}\big)_n.$$

Thus the occupied set at level n, when represented in binary sequence space, is

$$\mathcal{A}_n^\sigma = \Big\{x \in \Omega : \big(T_{\sigma(x)}\big)_n = 1\Big\}.$$

This representation makes the role of the Coin sequence explicit: it selects the particular tree automorphism through which the standard binary coordinates are read.

Equivalently, writing

$$r_{n(x)} = 2x_n - 1$$

for the digit-Rademacher functions, the membership signs admit the predictable-sign representation

$$2Y_{n(x)} - 1 = (-1)^{\varphi_n^{\sigma(x_1,\dots,x_{n-1})}} r_{n(x)}.$$

Thus the Rademacher description is simply another formulation of the same binary-tree recursion.

5.3 Conjugacy between different Coin sequences

The preceding theorem gives more than equality of measure-theoretic invariants.

Theorem 6 (Exact conjugacy). Let σ and τ be any two deterministic Coin sequences. Define

$$H_{\sigma,\tau} = T_\tau^{-1} \circ T_\sigma.$$

Then $H_{\sigma,\tau}$ is a measure-preserving isometry of the binary ultrametric space Ω, and for every level n,

$$H_{\sigma,\tau}(\mathcal{A}_n^\sigma) = \mathcal{A}_n^\tau.$$

Consequently, it also maps the limsup and liminf sets of one necklace onto those of the other.

Proof. By Theorem 5, both T_σ and T_τ are measure-preserving isometries of Ω. Their composition

$$H_{\sigma,\tau} = T_\tau^{-1} \circ T_\sigma$$

has the same properties.

For every $x \in \Omega$,

$$T_{\tau(H_{\sigma,\tau}(x))} = T_{\sigma(x)}.$$

Taking the n-th coordinate gives

$$\left(T_{\tau(H_{\sigma,\tau}(x))}\right)_n = \left(T_{\sigma(x)}\right)_n.$$

Therefore,

$$x \in \mathcal{A}_n^\sigma \Rightarrow H_{\sigma,\tau}(x) \in \mathcal{A}_n^\tau.$$

This proves the level-by-level identity. Since a bijection commutes with taking unions and intersections,

$$H_{\sigma,\tau}\left(\limsup_n \mathcal{A}_n^\sigma\right) = \limsup_n \mathcal{A}_n^\tau,$$

and

$$H_{\sigma,\tau}\left(\liminf_n \mathcal{A}_n^\sigma\right) = \liminf_n \mathcal{A}_n^\tau.$$

The theorem follows.

This result shows that all deterministic Coin sequences generate the same level-set structure up to an exact isometry of the binary ultrametric space. In particular, the Coin sequence does not produce distinct intrinsic binary-tree geometries.

5.4 An Explicit Interval Example

We now give a concrete example showing that the tree isometry constructed above need not preserve Euclidean distance.

Let $\sigma = (1, 1, 1, ...)$ and $\tau = (0, 0, 0, ...)$, and consider the conjugacy

$$H_{\sigma,\tau} = T_\tau^{-1} \circ T_\sigma.$$

For a binary sequence $\beta = (\beta_1, \beta_2, ...)$, write

$$P_{n-1}(\beta) = \left(\sum_{r=1}^{n-1} \beta_r\right) \bmod 2.$$

The triangular representation of the two tree maps gives

$$\varphi_n^{\sigma(\beta)} = P_{n-1}(\beta) \oplus (n \bmod 2),$$

and

$$\varphi_n^{\tau(\beta)} = 1 \oplus P_{n-1}(\beta).$$

Here $\oplus$ denotes addition modulo two.

Let $\gamma = H_{\sigma,\tau}(\beta)$. Since $T_{\tau(\gamma)} = T_{\sigma(\beta)}$, comparison of the n-th coordinates yields

$$\gamma_1 = \beta_1,$$

and, for $n \geq 2$,

$$\gamma_n = \beta_n \oplus P_{n-1}(\beta) \oplus ((n-1) \bmod 2) \oplus P_{n-1}(\gamma).$$

Consider the two non-dyadic points

$$x = 0.010101\ldots_2 = \frac{1}{3}$$

and

$$y = 0.101010\ldots_2 = \frac{2}{3}.$$

These are convenient non-dyadic points with simple periodic binary expansions.

Applying the preceding recursion gives

$$H_{\sigma,\tau}(x) = 0.00101010\ldots_2 = \frac{1}{6},$$

whereas

$$H_{\sigma,\tau}(y) = 0.11010101\ldots_2 = \frac{5}{6}.$$

Consequently,

$$|x - y| = \frac{1}{3},$$

but

$$\left|H_{\sigma,\tau}(x) - H_{\sigma,\tau}(y)\right| = \frac{2}{3}.$$

Thus $H_{\sigma,\tau}$ is not an isometry for the Euclidean metric.

This example does not contradict the tree-isometry property. The latter preserves the length of common binary prefixes, whereas Euclidean distance also depends on the binary tails following the common prefix.

5.5 What Changes under Interval Coding?

Let

$$\pi(\beta) = \sum_{n=1}^{\infty} \beta_n 2^{-n}$$

be the usual binary coding map from $\Omega = \{0,1\}^{\mathbb{N}}$ to $[0,1]$.

The map π is continuous and surjective, but it is not injective: dyadic rational numbers have two binary expansions.

Therefore, a tree automorphism on Ω does not automatically induce a well-defined map on the Euclidean interval. Such an induced map would have to send any two binary sequences representing the same real number to sequences representing the same real number.

The conjugacy in the preceding example illustrates this obstruction.

For the same choice

$$\sigma = (1, 1, 1, ...), \quad \tau = (0, 0, 0, ...),$$

the two sequences

$$\beta = 0.011111...{}_2$$

and

$$\beta' = 0.100000...{}_2$$

both represent the real number $\frac{1}{2}$.

However, the coordinate recursion gives

$$H_{\sigma,\tau}(\beta) = 0.000000...{}_2$$

and

$$H_{\sigma,\tau}(\beta') = 1.111111...{}_2.$$

Their images under π are respectively 0 and 1. Thus the two binary expansions of $\frac{1}{2}$ are not identified by this conjugacy.

In particular, this conjugacy does not descend to a single-valued map on the entire interval $[0, 1]$ under the usual binary coding.

On the set of points with unique binary expansions, one may still define a pointwise interval realization by applying the sequence-space map and then using π. The preceding example shows that this realization need not preserve Euclidean distance.

Accordingly, the tree conjugacy establishes an exact equivalence in binary sequence space, but it does not by itself establish a global homeomorphism, an order-preserving equivalence, or a bi-Lipschitz equivalence of the corresponding Euclidean realizations.

5.6 Equivalence questions

The results above allow us to distinguish several notions of equivalence.

1. **Binary-tree equivalence.** Two necklaces are equivalent if there is a rooted binary tree automorphism mapping every level set of one necklace onto the corresponding level set of the other. Theorem 6 shows that all Neruthes necklaces are equivalent in this sense.

2. **Measure-theoretic equivalence.** Two necklaces are equivalent if their level sets are mapped onto one another by a measure-preserving isomorphism, modulo null sets. Theorem 6 also gives this equivalence.

3. **Euclidean ambient equivalence.** Two interval realizations are equivalent if there is a homeomorphism of the interval mapping every level set of one realization onto the

corresponding level set of the other. This is a stronger question and is not settled by the binary-tree conjugacy.

4. **Euclidean metric equivalence.** One may further ask whether such a map can be chosen to be bi-Lipschitz, or whether the two realizations have the same quantitative Euclidean geometric invariants.

The last two notions are the appropriate setting for asking whether different Coin sequences can produce genuinely distinct interval geometries.

5.7 The classification problem

The exact binary-tree conjugacy reduces the classification problem. There is only one equivalence class under rooted binary tree isometries. Any nontrivial classification must therefore use structure that the binary ultrametric does not retain.

We propose the following questions.

1. For which pairs of Coin sequences does the binary-tree conjugacy descend to a well-defined homeomorphism of the Euclidean interval?
2. Can two Coin sequences produce necklaces that are not equivalent under an increasing homeomorphism of $[0, 1]$?
3. Are there Coin sequences whose interval realizations are not bi-Lipschitz equivalent?
4. How do finite-level quantities, such as the number and arrangement of occupied runs, depend on the Coin sequence?
5. For periodic or eventually periodic Coin sequences, can these Euclidean invariants be computed by finite-state methods?

These questions concern the interaction between a fixed binary-tree structure and its different realizations in the Euclidean interval. They provide a more precise formulation of the geometric problem suggested by the Neruthes necklace.

6 Finite-Stage Geometry and Special Coin Sequences

The measure of a finite-stage set records its total occupied length, but it does not record how that occupied region is distributed across the interval. A simple additional statistic is the number of connected components.

Since each A_n is a finite union of dyadic half-open intervals, its connected components are precisely its maximal consecutive runs of occupied cells. We denote their number by

$$R_n = \text{number of connected components of } A_n.$$

This statistic depends on the arrangement of the occupied cells, not merely on their number.

6.1 An Exact Component-Count Formula

At level n, write the parent pairs as

$$p_j = b_{2j}^{(n)} b_{2j+1}^{(n)}, \quad 0 \le j < 2^{n-1}.$$

Define their orientation indicators by

$$q_{n,j} = \begin{cases} 1 \text{ if } \ p_j = 10 \\ 0 \text{ if } \ p_j = 01. \end{cases}$$

Let

$$c_{n,j} = q_{n,j} \oplus \sigma_{2^{n-1}+j}.$$

By the definition of the Expand operation, the j-th parent pair produces the word 1001 exactly when $c_{n,j} = 1$, and produces 0110 exactly when $c_{n,j} = 0$.

Theorem 10 (Component-count formula). For every $n \geq 1$,

$$R_{n+1} = 2^{n-1} + \sum_{j=0}^{2^{n-1}-1} c_{n,j} - \sum_{j=0}^{2^{n-1}-2} c_{n,j}c_{n,j+1}.$$

Proof. Put $m = 2^{n-1}$, the number of parent pairs at level n.

A block of type 1001 contains two runs of ones, whereas a block of type 0110 contains one. If the blocks were separated by zeros, their total number of runs would therefore be

$$m + \sum_{j=0}^{m-1} c_{n,j}.$$

However, two consecutive blocks merge their boundary runs precisely when both blocks are of type 1001. Indeed, 1001 begins and ends with 1, whereas 0110 begins and ends with 0. Each adjacent pair of consecutive 1001 blocks therefore reduces the total run count by exactly one.

The number of such adjacent pairs is given by

$$\sum_{j=0}^{m-2} c_{n,j}c_{n,j+1}.$$

When $m = 1$, this sum is empty and evaluates to zero. Subtracting these mergers gives the stated formula. □

Corollary (Transition-count formula). Let $N_{10}(c_n)$ denote the number of occurrences of the adjacent pattern 10 in the binary word

$$c_n = \big(c_{n,0}, ..., c_{n,2^{n-1}-1}\big).$$

Then

$$R_{n+1} = 2^{n-1} + N_{10}(c_n) + c_{n,2^{n-1}-1}.$$

Proof. Since each $c_{n,j}$ is binary,

$$c_{n,j}\big(1 - c_{n,j+1}\big) = \begin{cases} 1 \\ \quad \big(c_{n,j}, c_{n,j+1}\big) = (1,0) \\ 0 \\ \quad \text{otherwise.} \end{cases}$$

Consequently,

$$\sum_{j=0}^{m-2} c_{n,j}\big(1 - c_{n,j+1}\big) = N_{10}(c_n).$$

Rewriting the component-count formula gives

$$R_{n+1} = m + \sum_{j=0}^{m-2} c_{n,j}\big(1 - c_{n,j+1}\big) + c_{n,m-1},$$

which proves the claim. □

The transition-count formula separates two contributions to the geometry of the next stage: the number of transitions from 1 to 0 in the orientation word and a terminal contribution determined by its final bit.

In particular, R_{n+1} is not determined by R_n alone. The formula depends on the local arrangement of the parent-pair orientations and the Coin values.

6.2 The Alternating Coin Sequence

Consider the alternating Coin sequence

$$\sigma = (1, 0, 1, 0, 1, 0, ...).$$

For this sequence, the component count admits an exact closed formula.

Theorem 11. For every $n \geq 2$, the stage-n configuration is

$$b^{(n)} = (0110)^{2^{n-2}},$$

where the right-hand side denotes the concatenation of 2^{n-2} copies of the word 0110.

Consequently, the number of connected components of A_n is

$$R_n = 2^{n-2}.$$

Proof.

At the first expansion, the initial word 10 is replaced using the first Coin bit, which is 1. Hence

$$b^{(2)} = 0110.$$

Suppose that, for some $n \geq 2$,

$$b^{(n)} = (0110)^{2^{n-2}}.$$

The adjacent parent pairs are therefore

$$01, 10, 01, 10, ...$$

Their orientations, with 01 encoded by 0 and 10 encoded by 1, form the alternating word

$$0, 1, 0, 1, ...$$

At level n, the Coin bits used are

$$\sigma_{2^{n-1}}, \sigma_{2^{n-1}+1}, ...$$

Since 2^{n-1} is even, this segment of the alternating Coin sequence is also

$$0, 1, 0, 1, ...$$

Thus the orientation of each parent pair agrees with its corresponding Coin bit. By the replacement rule, every parent pair produces the word 0110.

There are 2^{n-1} parent pairs, so

$$b^{(n+1)} = (0110)^{2^{n-1}}.$$

This completes the induction.

Finally, each copy of 0110 contains exactly one maximal run of consecutive 1s, and different copies are separated by zeros. Therefore the number of connected components is exactly the number of copies:

$$R_n = 2^{n-2}.$$

□

6.3 An Exact Formula for the Constant-One Coin Sequence

We now consider the constant-one Coin sequence

$$\sigma = (1, 1, 1, ...).$$

In this case, the component count admits an exact closed formula. The proof reveals a connection between the pair-orientation words of the Neruthes construction and the Thue–Morse sequence.

Let $t_j = s_2(j) \bmod 2$ denote the Thue–Morse sequence, where $s_2(j)$ is the sum of the binary digits of j.

Theorem 12. For the constant-one Coin sequence, the number of connected components of A_n is

$$R_n = \frac{2^n + (-1)^{n-1}}{3}$$

for every $n \geq 1$.

In particular,

$$R_n \sim \frac{2^n}{3}$$

as $n \to \infty$.

Proof.

For each stage n, let $q_{n,j}$ denote the orientation of the j-th parent pair:

$$q_{n,j} = \begin{cases} 1 \\ \quad \text{if the pair is } \ 10 \\ 0 \\ \quad \text{if the pair is } \ 01. \end{cases}$$

The initial configuration is $b^{(1)} = 10$, so $q_{1,0} = 1$.

Under a constant-one Coin sequence, a pair 01 produces 1001, while a pair 10 produces 0110.

The two consecutive pairs in the output therefore have orientations

$$\big(1 - q_{n,j}, q_{n,j}\big).$$

Consequently,

$$q_{n+1,2j} = 1 - q_{n,j},$$

$$q_{n+1,2j+1} = q_{n,j}.$$

Recall that the Thue–Morse sequence satisfies

$$t_{2j} = t_j, \quad t_{2j+1} = 1 - t_j.$$

Using the initial value $q_{1,0} = 1$, induction gives

$$q_{n,j} = t_j \oplus (n \bmod 2).$$

Since every Coin bit equals 1, the variables in the component-count formula satisfy

$$c_{n,j} = q_{n,j} \oplus 1 = t_j \oplus ((n+1) \bmod 2).$$

Thus, at each level, the word formed by the $c_{n,j}$ is either a Thue–Morse word or its bitwise complement.

We next count adjacent equal pairs in Thue–Morse words.

Let U_k be the finite word

$$U_k = t_0 t_1 ... t_{2^k - 1}.$$

Let a_k denote the number of occurrences of 11 as adjacent symbols in U_k, and let d_k denote the number of occurrences of 00.

The Thue–Morse recursion gives

$$U_{k+1} = U_k \overline{U_k},$$

where $\overline{U_k}$ is the bitwise complement of U_k.

The last symbol of U_k is $k \bmod 2$. Therefore,

$$a_{k+1} = a_k + d_k + \varepsilon_k,$$

$$d_{k+1} = d_k + a_k,$$

where $\varepsilon_k = 1$ if k is odd and $\varepsilon_k = 0$ if k is even.

Since $U_1 = 01$, we have

$$a_1 = d_1 = 0.$$

Induction using these recurrences yields

$$a_k = \begin{cases} \frac{2^{k-1}+1}{3} & k \text{ even} \\ \frac{2^{k-1}-1}{3} & k \text{ odd} \end{cases}$$

and

$$d_k = \begin{cases} \frac{2^{k-1}-2}{3} & k \text{ even} \\ \frac{2^{k-1}-1}{3} & k \text{ odd.} \end{cases}$$

We now apply the component-count identity from Theorem 10.

For a stage $N \geq 3$, put $k = N - 2$ and $m = 2^k$. The word c has length m. The Thue–Morse word is balanced, so both it and its complement contain exactly $\frac{m}{2}$ occurrences of 1.

If N is even, then k is even and c is the Thue–Morse word. Its number of adjacent 11 pairs is

$$a_k = \frac{2^{k-1}+1}{3}.$$

If N is odd, then k is odd and c is the complement of the Thue–Morse word. Its number of adjacent 11 pairs is therefore the number of adjacent 00 pairs in U_k, namely

$$d_k = \frac{2^{k-1}-1}{3}.$$

The component-count formula becomes

$$R_N = m + \frac{m}{2} - a_k$$

when N is even, and

$$R_N = m + \frac{m}{2} - d_k$$

when N is odd.

Substituting the corresponding expressions and using $m = 2^{N-2}$ gives

$$R_N = \begin{cases} \frac{2^N-1}{3} & N \text{ even} \\ \frac{2^N+1}{3} & N \text{ odd.} \end{cases}$$

Equivalently,

$$R_N = \frac{2^N + (-1)^{N-1}}{3}.$$

For $N = 1$, the initial configuration consists of the single interval corresponding to the word 10, so $R_1 = 1$, which also agrees with the formula.

For $N = 2$, the configuration is 0110, so $R_2 = 1$, which also agrees with the formula.

This completes the proof. □

6.4 The Thue–Morse Coin Sequence

For completeness, define the Thue–Morse sequence by

$$t_m = s_2(m) \bmod 2, \quad m \geq 0,$$

where $s_2(m)$ is the sum of the binary digits of m. We use the shifted indexing

$$\sigma_k = t_{k-1}, \quad k \geq 1.$$

This convention specifies the input Coin sequence only. The occupancy word at level n, the parent-pair orientation word, and the sequence of component counts are derived objects; none of them should be identified with the Thue–Morse sequence without a separate proof.

The component-count formula in Theorem 10 applies to this Coin sequence without modification. Its transition-count form expresses the component count in terms of the number of occurrences of the pattern 10 in the induced orientation word and the final bit of that word.

6.5 A Finite-Level Comparison

The component-count formula gives a direct way to compare different Coin sequences. Consider the following five choices:

1. The constant sequence $\sigma_k = 1$.
2. The constant sequence $\sigma_k = 0$.
3. The periodic sequence $\sigma = 101010....$
4. The periodic sequence $\sigma = 010101....$
5. The Thue–Morse sequence $\sigma_k = t_{k-1}$.

Starting from $b^{(1)} = 10$ and consuming the Coin sequence consecutively from left to right, the first ten component counts are:

Level	**All 1**	**All 0**	**10 periodic**	**01 periodic**	**Thue–Morse**
1	1	1	1	1	1
2	1	2	1	2	2
3	3	3	2	2	3
4	5	6	4	5	5
5	11	11	8	8	11
6	21	22	16	17	22
7	43	43	32	32	43
8	85	86	64	65	85
9	171	171	128	128	171
10	341	342	256	257	342

These values are exact finite computations. Additional computation extends the comparison through Level 31. They illustrate that the component count can distinguish some Coin sequences even though all of them have the same occupied measure at every level.

The values in the all-one column agree with the closed formula proved in Theorem 12. Likewise, the values in the alternating 1010... column agree with the exact formula established in Theorem 11.

The Thue–Morse example reveals a systematic pattern of coincidences. Its component counts agree with those of both constant Coin sequences at every odd level. At even levels, the agreement alternates between the two constant sequences: the Thue–Morse count agrees with the all-one count at levels congruent to 0 modulo 4, and with the all-zero count at levels congruent to 2 modulo 4.

The computations through Level 31 support a precise modulo-four formula for the Thue–Morse component count. We state the corresponding conjectures below. These finite computations provide evidence but do not establish the formulas for arbitrary levels.

6.6 The Induced Thue–Morse Orientation Words

The modulo-four relation concerns a statistic derived from the parent-pair orientation words, rather than the Thue–Morse Coin sequence directly.

Recall that

$$c_{n,j} = q_{n,j} \oplus \sigma_{2^{n-1}+j}.$$

For the shifted Thue–Morse Coin sequence, this becomes

$$c_{n,j} = q_{n,j} \oplus t_{2^{n-1}+j-1}.$$

Thus the input segment used at level n is a shifted block of the Thue–Morse sequence. The shift is important: the derived orientation word is not simply an ordinary Thue–Morse prefix.

By the transition-count formula,

$$R_{n+1} = 2^{n-1} + N_{10}(c_n) + c_{n,2^{n-1}-1}.$$

For the constant-one Coin sequence, the orientation words have a direct description in terms of the Thue–Morse sequence, as shown in the proof of Theorem 12. The corresponding component counts can therefore be evaluated through adjacent-pair statistics of finite Thue–Morse words.

For the Thue–Morse Coin sequence itself, however, the orientation words evolve under a different interaction between the orientation recursion and the shifted Coin input. The resulting words should not be identified with ordinary Thue–Morse prefixes without an independent derivation.

The Thue–Morse identities

$$t_{2m} = t_m, \quad t_{2m+1} = 1 - t_m$$

provide a possible starting point for deriving recursive relations for the induced orientation words. These identities alone do not establish the component-count formula: the evolution of the parent-pair orientation word must also be tracked.

A proof of the modulo-four conjecture would therefore require explicit control of these derived orientation words, or of the relevant adjacent-transition statistics. This provides a concrete combinatorial problem underlying the observed coincidence.

6.7 Component-Count Formulas and the Modulo-Four Pattern

We now establish the component-count formula for the constant-zero Coin sequence and record the observed modulo-four relation involving the Thue–Morse Coin sequence. The proofs use the component-count formula of Theorem 10 and the orientation-word recursion established in Theorem 12.

Lemma (Complement difference). Let $c = (c_0, ..., c_{m-1})$ be a binary word, and let $|\ c = (1 - c_0, ..., 1 - c_{m-1})$ be its bitwise complement. Define

$$F(c) = m + \sum_{j=0}^{m-1} c_j - \sum_{j=0}^{m-2} c_j c_{j+1}.$$

Then

$$F(|\ c) - F(c) = 1 - c_0 - c_{m-1}.$$

Proof. Expanding the definition gives

$$F(|\ c) - F(c) = \sum_{j=0}^{m-1} \big(1 - 2c_j\big) - \sum_{j=0}^{m-2} \big(1 - c_j - c_{j+1}\big).$$

The first sum equals $m - 2\sum_{j=0}^{m-1} c_j$, while the second equals

$$(m-1) - 2\sum_{j=0}^{m-1} c_j + c_0 + c_{m-1}.$$

Subtracting yields

$$F(|\ c) - F(c) = 1 - c_0 - c_{m-1}.$$

□

Proposition (Constant-zero component-count formula). For the constant-zero Coin sequence, the number of connected components of A_n is

$$R_n^{(0)} = \begin{cases} \frac{2^n+1}{3} & \\ \qquad n \text{ odd} & \\ \frac{2^n+2}{3} & \\ \qquad n \text{ even.} & \end{cases}$$

for every $n \geq 1$.

Proof. At level n, let $c^{(1)}$ denote the effective orientation word for the constant-one Coin sequence, and let $c^{(0)}$ denote the corresponding word for the constant-zero Coin sequence.

Since the two Coin sequences are bitwise complements, their effective orientation words are also bitwise complements.

For $n \geq 2$, put $k = n - 1$. By the orientation recursion in the proof of Theorem 12,

$$q_{k,j} = t_j \oplus (k \bmod 2),$$

where t_j is the Thue–Morse sequence. Consequently,

$$c_j^{(1)} = q_{k,j} \oplus 1 = t_j \oplus (n \bmod 2).$$

The first Thue–Morse bit is $t_0 = 0$, so

$$c_0^{(1)} = n \bmod 2.$$

The last index of the word is $2^{n-2} - 1$. Using the identity

$$t_{2^r-1} = r \bmod 2,$$

we obtain

$$c_{2^{n-2}-1}^{(1)} = 0.$$

Applying the complement-difference lemma to the two effective orientation words gives

$$R_n^{(0)} - R_n^{(1)} = 1 - (n \bmod 2).$$

Therefore,

$$R_n^{(0)} = \begin{cases} R_n^{(1)} & \\ & n \text{ odd} \\ R_n^{(1)} + 1 & \\ & n \text{ even.} \end{cases}$$

Substituting the constant-one formula from Theorem 12,

$$R_n^{(1)} = \frac{2^n + (-1)^{n-1}}{3},$$

gives the asserted expression. For $n = 1$, both sequences start from the same initial configuration, and the formula also holds.

□

Conjecture (Thue–Morse component-count formula). For the Thue–Morse Coin sequence, the number of connected components of A_n is

$$R_n^{\mathrm{TM}} = \frac{2^n + s_n}{3},$$

where

$$s_n = \begin{cases} 1 & \\ & n \text{ odd} \\ -1 & \\ & n \equiv 0 (\mathrm{mod}\, 4) \\ 2 & \\ & n \equiv 2 (\mathrm{mod}\, 4). \end{cases}$$

for every $n \geq 1$.

The formula is consistent with the finite computations through Level 31. In particular, it predicts the following exact coincidences with the two constant-sequence component-count formulas:

$$R_n^{\mathrm{TM}} = R_n^{(1)} = R_n^{(0)}$$

for every odd n, while

$$R_{4m}^{\mathrm{TM}} = R_{4m}^{(1)}$$

and

$$R_{4m+2}^{\mathrm{TM}} = R_{4m+2}^{(0)}$$

for every $m \geq 1$.

These identities explain the observed modulo-four pattern: the Thue–Morse component count agrees with the constant-one formula at levels divisible by four and with the constant-zero formula at levels congruent to two modulo four. At odd levels, all three component counts coincide.

The transition-count formula of Theorem 10 provides a useful framework for investigating this conjecture. Specifically, if c_n denotes the effective orientation word at level n, then

$$R_{n+1} = 2^{n-1} + N_{10}(c_n) + c_{n,2^{n-1}-1}.$$

Thus, proving the Thue–Morse formula requires controlling both the number of occurrences of 10 in the effective orientation word and its final bit.

The Thue–Morse identities

$$t_{2j} = t_j, \quad t_{2j+1} = 1 - t_j$$

provide recursive information about the Coin input, but they do not by themselves determine the adjacent-transition statistics of the resulting effective orientation words. In particular, the component-count formula cannot be established merely by comparing the effective words with ordinary Thue–Morse prefixes or by assuming a uniform adjacent-pair parity relation.

A complete proof of the conjecture therefore requires an additional recursive analysis of the effective orientation words, or an equivalent closed recurrence for their transition statistics. Until such a recurrence is established, the formula remains a conjecture rather than a theorem.

All the identities in this subsection concern component counts only. They do not imply equality of the finite-stage configurations, the parent-pair orientation words, or the underlying Coin sequences.

6.8 Scope of the Geometric Statistic

The component count is a useful first Euclidean invariant, but it is not a complete description of the geometry. It records how many connected pieces occur, while discarding their lengths, positions, and mutual distances.

Further statistics suggested by the construction include:

1. the number of occupied runs of each length;
2. the distribution of gaps between consecutive occupied components;
3. the largest occupied component and the largest vacant gap;
4. the ordered list of component lengths at each level.

These quantities may distinguish Coin sequences that have identical component counts. They are natural extensions of the finite-stage geometry of the Neruthes necklace.

7 Visualization

The following diagrams show the first nine stages of the Neruthes necklace. In each diagram, rows are ordered from the initial configuration to the finest displayed level. Black cells represent occupied dyadic intervals.

7.1 Constant Coin Sequences

7.1.1 All True



7.1.2 All False



7.2 Periodic Coin Sequences

7.2.1 Alternating 10



7.2.2 Alternating 01



7.3 A Non-periodic Coin Sequence

7.3.1 Thue–Morse



8 Conclusion and Open Problems

The Neruthes necklace is generated by a simple deterministic rule: each balanced parent pair is replaced by one of two balanced four-bit blocks, with the choice prescribed by a Coin sequence. The purpose of this article has been to investigate the mathematical structure produced by this recursion, rather than to regard it merely as an example of a probabilistic process.

The results reveal a structural contrast between symbolic and Euclidean geometry. At the symbolic level, the recursion has a universal structure: the associated rooted-tree automorphisms are conjugate, and the membership indicators have the same independent fair Bernoulli law for every deterministic Coin sequence. At the Euclidean level, however, binary coding does not preserve all of this equivalence. Interval liminf sets, finite-stage component counts, and covering dimensions can depend on the Coin sequence and on the conventions used at dyadic endpoints. We refer to this dependence of Euclidean properties on the binary coding, despite the underlying symbolic equivalence, as **coding sensitivity**. In this article, the term describes a structural phenomenon rather than a numerical invariant.

8.1 Established results

First, every finite stage has Lebesgue measure $\frac{1}{2}$. For every deterministic Coin sequence, the membership indicators form an independent sequence of fair Bernoulli random variables. Consequently, their empirical averages converge almost everywhere to $\frac{1}{2}$, while the limsup has full Lebesgue measure and the liminf has measure zero.

Second, the triangular bit transformation associated with each Coin sequence is a rooted binary tree automorphism. Any two such transformations are conjugate by a rooted-tree automorphism, which is also an isometry of the binary ultrametric space. Thus, the symbolic recursion has a common underlying tree structure, independent of the particular Coin sequence.

Third, the symbolic eventual-membership sets are countably infinite and dense. Their images under the binary coding map are also countably infinite and dense in $[0,1]$, with Hausdorff and packing dimensions zero and both lower and upper box dimensions equal to one. The interval liminf, however, requires additional care because dyadic rationals admit two binary expansions. In particular, for the constant-one Coin sequence, the interval liminf is empty, despite the existence of a countably infinite dense symbolic eventual-membership set. This example demonstrates that symbolic membership and interval membership need not agree at coding endpoints.

Fourth, the finite-stage component count provides a concrete Euclidean statistic for comparing different Coin sequences. For the alternating Coin sequence, we obtain

$$R_n = 2^{n-2}, \quad n \geq 2,$$

whereas for the constant-one Coin sequence,

$$R_n = \frac{2^n + (-1)^{n-1}}{3}.$$

The latter formula follows from an explicit connection between the pair-orientation words and the Thue–Morse sequence. In addition, finite computations reveal a modulo-four coincidence among the component counts associated with the constant-one, constant-zero, and Thue–Morse Coin sequences. This suggests that distinct deterministic inputs can produce systematically identical values of the same Euclidean statistic at infinitely many levels. We formulate this relation as a conjecture, rather than an established theorem, in the section on periodic and constant Coin sequences.

These results illustrate that identical symbolic tree structure does not imply identical finite-stage interval geometry, while also showing that the component count alone may fail to distinguish different Coin sequences.

Fifth, we construct a family of valid finite-state Coin sequences whose interval liminf sets have explicitly computable dimensions. For each integer $k \geq 2$, the corresponding interval liminf L_-^k is countable and satisfies

$$\dim_H L_-^k = \dim_P L_-^k = 0,$$

while its lower and upper box dimensions coincide and are given by

$$\underline{\dim}_B L_-^k = \overline{\dim}_B L_-^k = d_k = \log_2(\lambda_k).$$

Here, λ_k is the unique real root greater than one of

$$\lambda_k^k = \lambda_k^{k-1} + ... + \lambda_k + 1.$$

The closure $K_k = \overline{L_-^k}$ is the binary projection of the subshift obtained by forbidding k consecutive zeros, and it satisfies

$$\dim_H K_k = \dim_B K_k = d_k.$$

The dimensions d_k are strictly increasing and converge to one. For the same family, the interval limsup is $[0, 1)$. In particular, this construction gives a sequence of examples in which a countable interval liminf has positive box dimension, while its closure has positive Hausdorff dimension.

Together, these results establish that the balanced recursion has a universal symbolic probabilistic and tree structure, but can exhibit substantially different Euclidean behaviour under different Coin sequences.

8.2 Open Problems

The contrast between symbolic equivalence and Euclidean dependence leads to several questions. The following problems are intended to identify the structures that govern this transition.

1. **Finite-state dimension formulae.**

The finite-state examples above suggest that the box dimension of an interval liminf may be determined by the exponential growth rate of an associated language of admissible binary prefixes. Can this principle be established for a general class of finite-state Coin sequences?

More precisely, given a finite-state description satisfying the sibling-complement condition required by the balanced Expand rule, one may construct a finite directed graph whose paths encode admissible prefixes. Under what conditions does the growth rate of these paths determine the lower and upper box dimensions of the corresponding interval liminf and its closure? When the admissible paths are described by a single finite nonnegative transition matrix, can the dimension always be expressed through its Perron-Frobenius spectral radius? A further objective is to characterize when the liminf itself has zero Hausdorff dimension while its closure has positive Hausdorff dimension.

More generally, for parameterized families of finite-state Coin systems, how does the dimension depend on the automaton parameters? Can one identify parameter values at which the dimension law changes its analytic form or the underlying admissible-path structure changes?

2. **Finite-state realization of symbolic constraints.**

The family constructed in this article is based on a particularly simple forbidden-word condition. A broader question is to determine which symbolic constraints can arise from valid finite-state Coin sequences.

Which regular prefix languages and finite labeled graph path spaces can occur as symbolic closures of interval liminf sets generated by the balanced Expand rule? In the shift-invariant case, which subshifts of finite type or sofic shifts can be realized? The sibling-complement condition imposes a restriction on the transitions: for every state q, the two successor states must have complementary output bits. It is therefore not clear whether arbitrary finite-state symbolic constraints can be represented within the Neruthes construction, or whether this local balance condition defines a substantially smaller class.

3. **Component-count laws and inverse problems.**

The exact component-count formulas obtained for the alternating and constant-one Coin sequences raise two related questions.

Can one derive exact recurrences or asymptotic growth laws for the component counts of other periodic and automatic Coin sequences, including the Thue–Morse Coin sequence? In particular, when the Coin sequence is generated by a finite automaton, does the component-count sequence admit a finite-dimensional matrix representation or another effective algebraic description?

There is also an inverse problem. Does the complete component-count sequence determine the Coin sequence up to a natural equivalence, or can one construct inequivalent Coin sequences with identical component-count sequences?

The modulo-four coincidence conjectured for the constant-one, constant-zero, and Thue–Morse Coin sequences provides a concrete starting point. It predicts that distinct Coin sequences can have identical component counts at infinitely many levels, although this does not imply equality of their complete component-count sequences.

A stronger question is whether two inequivalent Coin sequences can have identical component-count sequences at every level. More generally, one may ask whether the number of levels at which two component-count sequences agree can be characterized in terms of the symbolic complexity or automaton structure of the corresponding Coin sequences.

These questions would clarify how much information about the symbolic recursion is retained by this finite-stage Euclidean statistic.

4. **Regularity of the interval realization.**

The following formulation makes precise what it means for a symbolic tree conjugacy to descend to the Euclidean interval.

The conjugacy between two symbolic tree automorphisms does not automatically induce a well-defined continuous map of the Euclidean interval. To formulate this issue precisely, let

$$\Omega_{\mathrm{can}} = \{\beta \in \{0,1\}^N : \text{beta is not eventually } 1\}$$

be the canonical binary coding space, and let T_σ denote the tree automorphism associated with a Coin sequence σ. Define

$$D_\sigma = \Big\{\beta \in \Omega_{\mathrm{can}} : T_{\sigma(\beta)} \in \Omega_{\mathrm{can}}\Big\}.$$

On the coded set $\pi(D_\sigma)$, the induced interval map is

$$S_{\sigma(\pi(\beta))} = \pi\Big(T_{\sigma(\beta)}\Big).$$

For which pairs of Coin sequences is this partial map well-defined and continuous? Under what additional conditions does it admit a continuous extension to $[0,1]$, or an extension that is monotone, absolutely continuous, or quantitatively regular? In particular, can one characterize continuity and the existence of one-sided limits at dyadic rationals directly from the corresponding tree automorphisms and Coin sequences?

5. **Euclidean equivalence of symbolic closures.**

For a Coin sequence σ, write

$$K_\sigma = \overline{L_-^\sigma}$$

for the closure of its interval liminf set. Since the interval liminf itself may be countable even when its closure has positive dimension, geometric equivalence questions should distinguish these two objects.

Which geometric properties of K_σ are invariant under homeomorphism, bi-Lipschitz equivalence, or quasisymmetric equivalence? Can the automaton structure or the symbolic constraint defining K_σ be recovered from any of these geometric invariants? More specifically, can two non-equivalent finite-state Coin systems generate closures that are bi-Lipschitz equivalent, or does the geometry retain enough information to distinguish their underlying symbolic constraints?

A further question concerns Fourier dimension. For finite-state examples such as the family K_k, do the Hausdorff and Fourier dimensions coincide? If not, can the Fourier-dimension deficit $\dim_H K_\sigma - \dim_F K_\sigma$ be characterized or bounded in terms of the automaton structure and the arithmetic properties of the associated coding?

6. **The box-dimension spectrum.**

The preceding problem concerns the existence and form of spectral dimension formulae. The present problem concerns the range of dimension values that can be realized by

varying the Coin sequence. The family d_k provides a first example of a monotone dimension law governed by a Perron root. It remains to determine whether more general finite-state families exhibit qualitatively different parameter-dependent dimension laws, including genuine critical phenomena.

The family L_-^k realizes the dimensions $d_k = \log_2(\lambda_k)$, where d_k increases to one. This gives infinitely many distinct values in $(0, 1)$, but does not determine the full set of dimensions attainable under the balanced Expand rule.

Which values can occur as lower and upper box dimensions of interval liminf sets generated by deterministic Coin sequences? Can one construct examples for which the lower and upper box dimensions are different? More generally, is there a useful characterization of the attainable dimension pairs

$$\left(\underline{\dim}_B L_-^\sigma, \overline{\dim}_B L_-^\sigma\right)$$

in terms of the symbolic complexity or automaton structure of σ?

8.3 A Longer-Term Interface

A possible longer-term direction is to investigate whether automaton-constrained covering constructions can provide tractable models for certain arithmetic covering problems, including restricted Kakeya-type settings. Such a connection would require an explicit arithmetic model and quantitative estimates linking symbolic complexity to the relevant covering or additive structure. At present, this should be regarded as a research direction rather than an established connection.

8.4 Final perspective

The Neruthes necklace exhibits two complementary forms of structure. Its balanced local recursion gives rise to a universal symbolic probabilistic law and a universal rooted-tree equivalence. Its Euclidean realization, in contrast, retains information about the Coin sequence through endpoint conventions, component counts, and fractal dimensions.

The finite-state constructions provide a concrete setting in which this distinction can be studied exactly. They show that a simple local balance constraint can generate countable exceptional sets with nontrivial box dimension and closures governed by finite-state symbolic dynamics. This suggests that the next step is not merely to produce further examples, but to identify general principles connecting automaton structure, admissible-prefix growth, and Euclidean dimension.

The central problem is therefore to understand precisely which features of the symbolic recursion survive interval coding, which are altered by the coding map, and how the resulting Euclidean geometry can be classified.

9 Acknowledgment

The author thanks ChatGPT and DeepSeek for discussions that helped clarify the distinction between the measure-theoretic structure and the Euclidean realization of the Neruthes necklace, and the Typst developers for the typesetting system.