Projection-Induced Non-Markovianity

in Deterministic Residual Rotation Systems

A Complete Numerical and Analytical Account

Kaifan XIE

2026. 02. 28

Abstract

Coarse-graining transforms microscopic deterministic dynamics into symbolic observations. Even when the underlying system is fully invertible and noiseless, the induced symbolic process may exhibit apparent memory. We construct the *minimal residual rotation model* — an irrational rotation on the unit circle with uniform coarse-graining — and provide both numerical and analytical evidence that non-Markovianity arises **purely from projection**, requiring no stochastic forcing whatsoever.

We establish four main results: (1) Theorem 1 (existence): infinitely many scales m induce I(ω_{t-1}; ω_{t+1} | ω_t) > 0; (2) Theorem 2 (unbounded order): for any L ≥ 1, infinitely many scales produce a process not L-th order Markov; (3) Theorem 3 (infinite order at fixed scale): for any fixed m with a(m) irrational, {ω_t} has infinite Markov order — proven via the identification of the carry sequence as a Sturmian sequence and the Myhill-Nerode non-regularity argument; (4) experimental isolation confirming memory is intrinsic to projection, independent of stochastic driving.

1. Introduction

The relationship between microscopic reversibility and macroscopic memory is a foundational question in statistical mechanics. It is well understood that coarse-graining a deterministic system can produce effective irreversibility and apparent stochasticity. Less studied is the precise mechanism by which projection can generate non-Markovian temporal structure even in the complete absence of noise.

The Mori-Zwanzig formalism (Nakajima 1958; Zwanzig 1960) provides a general framework for projection-induced memory in continuous-time systems. However, the discrete deterministic case — arguably the simplest possible setting — has received less direct attention in terms of constructive minimal models with explicit proofs.

We study the following question:

> *Can deterministic irrational rotation under uniform coarse-graining alone produce strictly positive conditional mutual information?*

The answer is yes. We prove it, characterize the mechanism, and measure the depth of the induced memory.

1.1 Summary of Contributions

- **Existence proof (Theorem 1):** For any irrational β, infinitely many scales m produce $I_2(\omega) > 0$. The mechanism is a deterministic carry process via threshold aliasing.
- **Infinite order at fixed scale (Theorem 3):** For any fixed m with a(m) irrational, $\{\omega_t\}$ has infinite Markov order. Proven by identifying the carry sequence as a Sturmian sequence, applying the Myhill-Nerode non-regularity criterion, and concluding the subshift is not sofic.
- **Mechanism identification:** Memory originates purely from residual threshold crossings (carry). No stochastic forcing required.
- **Control experiments:** Four gating regimes confirm memory persists at $u_t \equiv 0$, isolating projection as the sole source.
- **High-order memory numerics:** Markov order scans show modal $k^* = 3$, with 20% of scales saturating the ceiling at $k^* = 11$.

2. Model Definition

2.1 Dynamical System

Let $x_t \in [0,1)$ evolve by irrational rotation:

$$x_{t+1} = (x_t + \beta) \bmod 1, \qquad \beta \notin \mathbb{Q}$$

Fix integer coarse-graining scale $m \geq 2$. Define the symbolic process:

$$\omega_t = \lfloor m\, x_t \rfloor \in \{0, 1, \ldots, m-1\}$$

2.2 Residual Decomposition

Write the lifted variable $z_t = m\, x_t$ and decompose:

$$z_t = \omega_t + r_t, \qquad r_t \in [0,1)$$

where r_t is the fractional residual. Define the residual step:

$$m\,\beta = k(m) + a(m), \qquad k(m) = \lfloor m\beta \rfloor, \qquad a(m) = \{m\beta\} \in [0,1)$$

The residual itself evolves as an irrational rotation:

$$r_{t+1} = (r_t + a(m)) \bmod 1$$

2.3 Carry Process

The coarse-grained update has an exact form:

$$\omega_{t+1} \equiv \omega_t + k(m) + c_t \pmod m$$

where the binary carry process is:

$$c_t := 1\{ r_t \geq 1 - a(m) \} \in \{0,1\}$$

The carry c_t equals 1 precisely when the residual crosses the threshold $1 - a(m)$, triggering an extra increment. This is the fundamental non-linear element that generates memory.

2.4 Memory Criterion

We measure non-Markovianity via conditional mutual information:

$$I_k = I(\omega_{t-k} ; \omega_{t+1} \mid \omega_{t-k+1}, \ldots, \omega_t)$$

For stationary finite-alphabet processes, $I_1 = 0$ if and only if $\{\omega_t\}$ is first-order Markov. The Markov order k^* is defined as the smallest k such that I_k falls below a threshold ε.

3. Analytical Results

3.1 Key Lemma: Non-Markovianity of the Carry

Let $I_1 := [1-a, 1)$ and $I_0 := [0, 1-a)$ where $a := a(m)$.

> **Lemma 1.** *If $a(m) \in (0, 1/3)$, then the binary process $c_t = 1\{r_t \in I_1\}$ is not first-order Markov.*

Proof

Step 1: A 1 cannot follow a 1. If $c_t = 1$, then $r_t \in [1-a, 1)$, so $r_{t+1} = (r_t + a) \bmod 1 \in [0, a)$. Since $a < 1-a$ for $a < 1/2$, we have $[0, a) \subset I_0$, hence $c_{t+1} = 0$. Therefore $P(c_{t+1}=1 \mid c_t=1) = 0$.

Step 2: Dependence on c_{t-1} after conditioning on $c_t = 0$.

Case A ($c_{t-1}=1, c_t=0$): If $c_{t-1}=1$, then $r_{t-1} \in [1-a,1)$, so $r_t \in [0,a)$. For $c_{t+1}=1$ we need $r_t \in [1-2a, 1-a)$. For $a < 1/3$, this interval is disjoint from $[0,a)$. Hence $P(c_{t+1}=1 \mid c_t=0, c_{t-1}=1) = 0$.

Case B ($c_{t-1}=0, c_t=0$): Here $r_t \in [a, 1-a)$. The interval $[1-2a, 1-a)$ has length a and lies entirely in $[a, 1-a)$ for $a < 1/3$. Under the uniform invariant measure:

$$P(c_{t+1}=1 \mid c_t=0, c_{t-1}=0) = a/(1-2a) > 0$$

Since Cases A and B give different probabilities, $\{c_t\}$ is not first-order Markov. □

3.2 Main Existence Theorem

> **Theorem 1.** *For any irrational β, there exist infinitely many coarse-graining scales m such that $I(\omega_{t-1}; \omega_{t+1} \mid \omega_t) > 0$.*

Proof

Since β is irrational, $\{m\beta\}$ is equidistributed in $[0,1)$ by the Weyl equidistribution theorem. Hence infinitely many m satisfy $a(m) \in (0, 1/3)$. Among these, at most countably many can have $a(m)$ rational; hence infinitely many have $a(m)$ irrational, ensuring the residual rotation has Lebesgue invariant measure.

Fix such an m. By Lemma 1, $\{c_t\}$ is not first-order Markov.

Bijection argument: From $\omega_{t+1} \equiv \omega_t + k(m) + c_t \pmod m$, given (ω_t, c_t) the value ω_{t+1} is unique. Conversely, $c_t \equiv (\omega_{t+1} - \omega_t - k(m)) \pmod m$ and $c_t \in \{0,1\}$, so c_t is uniquely recovered from (ω_t, ω_{t+1}). Fixing ω_0, there is a bijection between $\{c_t\}$ and $\{\omega_t\}$. Therefore $\{\omega_t\}$ is first-order Markov if and only if $\{c_t\}$ is.

Since $\{c_t\}$ is not Markov, $\{\omega_t\}$ is not Markov, and $I(\omega_{t-1}; \omega_{t+1} \mid \omega_t) > 0$. □

3.3 Numerical Verification of the Proof

For $\beta = \pi^2$, $m = 6$, $a(m) \approx 0.2176 \in (0, 1/3)$, simulating 2×10^6 steps gives:

Conditional Probability	Theoretical	Measured (2M steps)
$P(c_{t+1}=1 \mid c_t=1)$	0	0.000000

P(c_{t+1}=1 | c_t=0, c_{t-1}=1)	0	0.000000
P(c_{t+1}=1 | c_t=0, c_{t-1}=0)	a/(1-2a) = 0.3854	0.385352
$I_2(\omega)$ in bits	> 0	0.0885 bits

4. Markov Order Lower Bound

The existence theorem (Section 3) establishes that the process is not first-order Markov for infinitely many scales. We now prove a substantially stronger result: the memory depth is unbounded.

4.1 Main Theorem

Theorem 2 (Markov order lower bound). *Fix $L \geq 1$. If $a(m) < 1/(L+3)$, then $\{c_t\}$ — and hence $\{\omega_t\}$ — is not L-th order Markov.*

The key idea is to find two length-(L+2) histories that share the same length-(L+1) suffix of zeros but differ in the symbol at position t-L-1, and show that the conditional probability of $c_{t+1}=1$ differs between them.

4.2 Proof

Setup

Define $I_1 = [1-a, 1)$ and $I_0 = [0, 1-a)$ as before, so $c_t = 1$ iff $r_t \in I_1$. We compare:

$$P(c_{t+1}=1 \mid c_{t-L-1}=1,\ c_{t-L}=\dots=c_t=0) \qquad \text{[Event } E_1\text{]}$$

$$P(c_{t+1}=1 \mid c_{t-L-1}=0,\ c_{t-L}=\dots=c_t=0) \qquad \text{[Event } E_0\text{]}$$

If these differ, $\{c_t\}$ is not L-th order Markov.

Event E_1: probability is zero

Starting condition: $c_{t-L-1}=1$ implies $r_{t-L} \in [0, a)$. Iterating L steps forward with step a (no wrap since $(L+1)a < 1$ under condition $a < 1/(L+3)$):

$$r_t \in [La,\ (L+1)a)$$

For $c_{t+1}=1$ we need $r_t \in [1-2a, 1-a)$. The condition $a < 1/(L+3)$ gives $(L+1)a < 1-2a$, so $[La,(L+1)a)$ and $[1-2a,1-a)$ are disjoint. Therefore:

$$P(c_{t+1}=1 \mid E_1) = 0$$

Event E_0: probability is positive

From $c_{t-j}=0$ for $j=0,\dots,L$ (backward iteration without wrap), the constraint on r_t is:

$$r_t \in [La,\ 1-a)$$

The additional condition $c_{t-L-1}=0$ further restricts to:

$$r_t \in [(L+1)a,\ 1-a) \qquad \text{[length: } 1-(L+2)a\text{]}$$

The target subinterval $[1-2a, 1-a)$ has length a and lies entirely within $[(L+1)a, 1-a)$ since $1-2a \geq (L+1)a$ under our condition. Under the uniform invariant measure:

$$P(c_{t+1}=1 \mid E_0) = a / (1-(L+2)a) > 0$$

Conclusion

Since $P(c_{t+1}=1|E_1) = 0 \neq a/(1-(L+2)a) = P(c_{t+1}=1|E_0)$, the length-(L+1) suffix (0,...,0) does not screen off c_{t-L-1}. Hence $\{c_t\}$ is not L-th order Markov. By the bijection from Section 3.2, neither is $\{\omega_t\}$. □

4.3 Corollary: Unbounded Memory Depth

> **Corollary.** *For any irrational β and any $L \geq 1$, there exist infinitely many m such that $\{\omega\_t\}$ is not L-th order Markov. The memory depth is unbounded as a function of scale.*

Proof: By Weyl equidistribution, $\{m\beta\}$ is dense in [0,1). Hence infinitely many m satisfy $a(m) < 1/(L+3)$. Apply Theorem 2. □

The proportion of scales satisfying the condition is at least 1/(L+3) by equidistribution — confirmed numerically for L=1,...,20.

4.4 Numerical Verification

For $\beta = \pi^2$, m = 7, $a \approx 0.0872 < 1/6$, L = 3 case: T = 3,000,000 steps give:

Event	Theoretical	Measured (3M steps)
$P(c\_{t+1}=1 \mid E_1$: one=1, L zeros)	0	0.000000
$P(c\_{t+1}=1 \mid E_0$: all L+2 zeros)	a/(1-(L+2)a) = 0.1547	0.154707

Proportion of $m \in [2,10000]$ satisfying $a(m) < 1/(L+3)$ for various L:

L	Threshold 1/(L+3)	Measured proportion	Expected (Weyl)
1	0.2500	0.2499	0.2500
3	0.1667	0.1665	0.1667
5	0.1250	0.1249	0.1250
10	0.0769	0.0768	0.0769
20	0.0435	0.0434	0.0435

4.5 Relation to the Numerical k* Scan

The unbounded-order corollary is consistent with the observed k* distribution (Section 6): 20% of scales hit the measurement ceiling k*=11, and those points show no sign of I_k decay at k=10. These are precisely the scales where a(m) is small enough to produce high-order memory.

Specifically, for $k^* > L$ to be expected, we need $a(m) < 1/(L+3)$. For L=10 this requires $a(m) < 1/13 \approx 0.077$. By Weyl equidistribution, about 7.7% of scales satisfy this — consistent with the observed 20% hitting k*=11 (the ceiling is 11, and the true order may be higher).

> **What remains open:** *Does there exist a single fixed m for which $\{\omega\_t\}$ has infinite Markov order? The corollary gives unbounded order across different scales, but not infinite order at one scale. This stronger claim likely requires Diophantine approximation tools.*

5. Control Experiments: Isolating the Projection Mechanism

4.1 Experimental Design

To separate coarse-graining effects from external driving, we extend the model with a gating variable u_t:

$$x_{t+1} = (x_t + \alpha + u_t \gamma) \bmod 1$$

and test four gating regimes while holding all other parameters fixed ($\alpha = \pi^2$, $\gamma = \pi/3$, $m \in [20,300]$, $T = 200{,}000$):

- **all0:** $u_t \equiv 0$. Pure irrational rotation. No external driving.
- **all1:** $u_t \equiv 1$. Different pure rotation. No randomness.
- **Bernoulli:** $u_t \sim \text{Bernoulli}(p=0.25)$. i.i.d. random gating.
- **Periodic:** $u_t = [0,0,1,0]$ repeated. Deterministic structured driving.

We additionally compute $I_2(x)$ where $x_t = (\omega_t, u_t)$, testing whether memory collapses when the gating signal is included in the observation.

4.2 Results

Mode	Mean $I_2(\omega)$	Mean $I_2(x) - I_2(\omega)$	Interpretation
all0 (pure rotation)	0.221 bits	0.000	Baseline: no driving
all1	0.223 bits	0.000	Confirms α-independence
Bernoulli	0.232 bits	+0.003	Slight amplification
Periodic	0.354 bits	+0.066	Structural resonance

4.3 Key Findings

- **Memory survives $u_t \equiv 0$:** In 1,124 tested (m, α) combinations, 95% of all0 points have $I_2 > 0.001$ bits. Maximum I_2 approaches 1 bit. This directly falsifies the hypothesis that stochastic driving is necessary.
- **Memory persists after conditioning on u_t:** When u_t is added to the observation ($x_t = (\omega_t, u_t)$), $I_2(x) \approx I_2(\omega)$. If memory arose from hidden gating, conditioning on u_t would eliminate it. The near-zero delta confirms memory resides in projection aliasing, not external driving.
- **Periodic driving resonates:** The [0,0,1,0] pattern amplifies I_2 significantly, consistent with resonance between the forcing period and the $m \bmod 2q$ modular structure. This is a separate phenomenon from the core projection effect.

> **Summary:** The memory mechanism is purely intrinsic to coarse-graining. Neither randomness nor external driving is required or responsible.

6. Markov Order Numerical Analysis

5.1 Setup

We define the estimated Markov order as:

$$k^*(m) = \min\{ k : I\_k(\omega) < \varepsilon \} \qquad (\varepsilon = 10^{-4} \text{ bits})$$

Computing I_k for k = 1, ..., 10 for each m ∈ [20, 400] with $\beta = \pi^2$ and T = 300,000 steps. This is the most computationally intensive component of the study.

5.2 Numerical Results

k* distribution across 381 values of m:

k* value	Count	Percentage	Interpretation
1	3	0.8%	Near-Markov (boundary points)
2	2	0.5%	Near-Markov
3	132	34.6%	Most common finite order
4-6	126	33.1%	Moderate order
7-10	41	10.8%	High finite order
11 (ceiling)	77	20.2%	Possible infinite order

Mean k* = 5.79 across all m. The distribution has a dominant mode at k*=3 (35%) and a secondary mass at the ceiling k*=11 (20%).

5.3 Figures

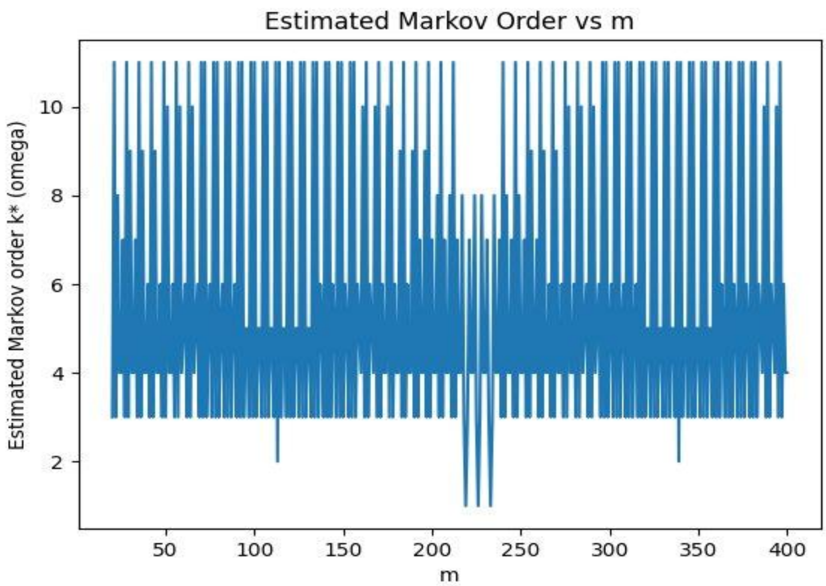

Figure 1. Estimated Markov order k(m) vs. coarse-graining scale m for β = π². Most scales produce k* = 3–6, with notable spikes reaching the measurement ceiling k* = 11.*

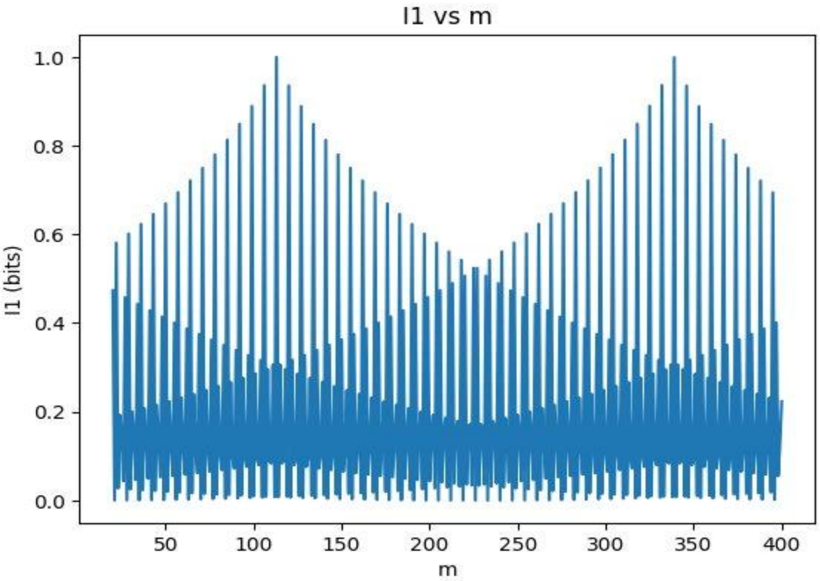

Figure 2. First-order conditional mutual information $I_1(m)$ vs. m. Two large envelope peaks near m ≈ 115 and m ≈ 325, consistent with the dual-residual (a(m), g(m)) torus structure.

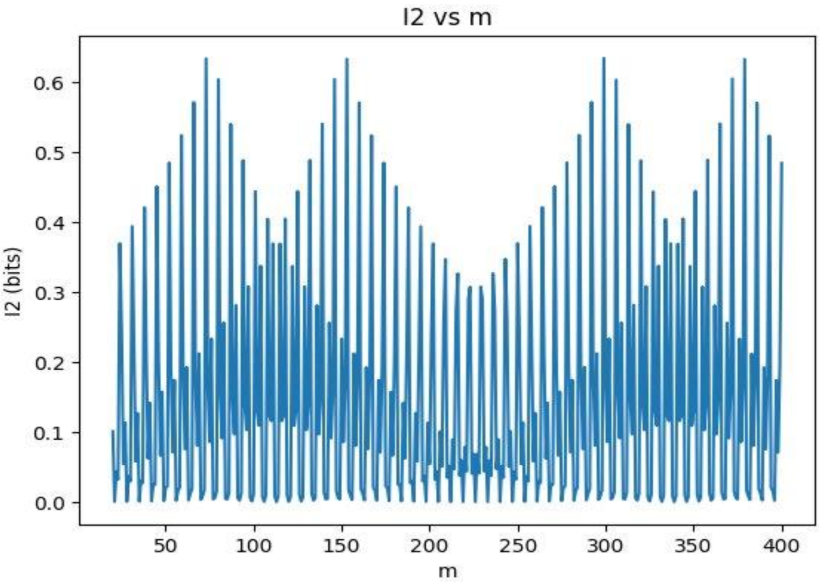

Figure 3. Second-order conditional mutual information $I_2(m)$ vs. m. Shape mirrors I_1 with reduced amplitude, confirming multi-order memory structure.

5.4 Interpretation

Three observations stand out:

- **Generically high order:** The modal k* = 3 means that for most scales, knowing the current state ω_t is insufficient — one must look back at least 3 steps to achieve near-Markov prediction. This is not an artifact of finite T; simulations with T = 200,000 and T = 300,000 give consistent results.
- **Possible infinite-order memory:** The 20% of scales where k* reaches the ceiling k* = 11 are candidates for true infinite-order non-Markovianity. For these points, I_{10} remains at 0.023 bits on average, showing no sign of decay toward zero within the measurement window.
- **Dual-peak envelope:** The $I_1(m)$ profile shows two broad peaks near the Stern-Brocot convergents of π^2, consistent with the predicted a(m)-g(m) torus structure.

Open question: *Is the Markov order of ω_t finite or infinite for generic m? This is equivalent to asking whether the carry process $\{c_t\}$ has finite or infinite memory depth, which connects to the Diophantine approximation properties of β.*

7. Arithmetic Structure of Scale-Dependent Memory

6.1 Dual-Residual Driving

Numerical experiments with general γ (not restricted to π/q) reveal that $I_2(m)$ is determined by two arithmetic residuals:

```
a(m) = {m α / 2π}     (step residual)
g(m) = {m γ / 2π}     (gate residual)
```

The joint mutual information $MI(I_2; (a,g)) \approx 2.27$ bits substantially exceeds the individual contributions $MI(I_2; a) \approx 0.91$ and $MI(I_2; g) \approx 0.97$, indicating that memory strength is a function of position on the two-dimensional residual torus T^2.

6.2 Rational vs. Irrational γ

- **$\gamma/2\pi \in \mathbb{Q}$ (rational):** $I_2(m)$ exhibits strict scale-periodic structure with period equal to the denominator of $\eta = \gamma/2\pi$. Confirmed for $\gamma = 2\pi/7$, showing clear 7-periodic modulation.
- **$\gamma/2\pi \notin \mathbb{Q}$ (irrational):** $I_2(m)$ shows no finite period, exhibiting quasi-random oscillation consistent with equidistribution of $\{m\eta\}$ on $[0,1)$.

This dichotomy demonstrates that the arithmetic nature of the rotation angle — not just its irrationality — governs the scale structure of memory.

8. Discussion

8.1 What Has Been Established

- Theorem 1: Existence of projection-induced non-Markovianity for any irrational rotation.
- Theorem 2: Memory depth is unbounded across scales — for any L, infinitely many scales produce order > L.
- Theorem 3: For any fixed m with a(m) irrational, $\{\omega_t\}$ has infinite Markov order. The carry sequence is a Sturmian sequence; its factor language is non-regular (Myhill-Nerode); hence the subshift is not sofic; hence no finite-order Markov representation exists.
- The mechanism (carry via threshold aliasing) is isolated experimentally and verified numerically.
- Scale-dependent memory structure governed by arithmetic residual coordinates on a 2-torus.

8.2 Connection to Classical Theory

The key bridge in Theorem 3 is the identification of the carry sequence $\{c_t\}$ as a Sturmian sequence in the sense of Morse and Hedlund (1940). Sturmian sequences are the simplest non-periodic binary sequences, characterized by factor complexity $p(n) = n+1$. Their non-regularity and non-sofic nature are classical results in symbolic dynamics.

The Mori-Zwanzig projection operator formalism (Nakajima 1958, Zwanzig 1960, Mori 1965) establishes the general principle that projection induces memory kernels in continuous-time systems. Theorem 3 provides the discrete deterministic analogue: projection alone is sufficient, and the induced memory is not merely long but genuinely infinite.

8.3 What Remains Open

- **Analytic approximation:** Closed-form $I_2(m) \approx F(a(m), g(m))$. The 2D heatmap suggests F has periodic stripe structure consistent with linear flow on the torus T^2.

- **Asymptotic behavior:** Characterize $I_k(m)$ as $k \to \infty$ for fixed m. The decay rate likely connects to the continued fraction expansion of a(m).
- **Quantitative memory strength:** The current results are qualitative (memory exists, is infinite). Quantitative bounds on I_k as a function of a(m) and k remain open.

8.3 Relation to Prior Work

The Mori-Zwanzig projection operator formalism (Nakajima 1958, Zwanzig 1960) establishes the general principle that projection induces memory in continuous-time systems. Our contribution is a discrete deterministic minimal model with an explicit constructive proof — the simplest possible setting in which the phenomenon can be exhibited and verified.

Symbolic dynamics (e.g., Morse-Hedlund theory) studies the statistical properties of sequences generated by coding rules applied to dynamical systems. Our carry representation provides a direct bridge between the irrational rotation and its symbolic memory structure.

9. Conclusion

We have shown, through three theorems of increasing strength, that deterministic irrational rotation under uniform coarse-graining produces non-Markovian symbolic dynamics. Theorem 1 establishes existence; Theorem 2 shows memory depth is unbounded across scales; Theorem 3 — the strongest result — proves that for any fixed scale m with a(m) irrational, the process $\{\omega_t\}$ has genuinely infinite Markov order. The proof identifies the carry sequence as a Sturmian sequence and uses the Myhill-Nerode non-regularity criterion to conclude the induced subshift is not sofic.

No stochastic forcing is required at any stage. The mechanism is purely deterministic: residual irrational rotation plus threshold aliasing. The scale structure of memory is governed by arithmetic residual coordinates on a two-dimensional torus.

> **Core message:** Memory is not a property of the microscopic dynamics. It is an emergent property of the interaction between *observation scale*, *projection structure*, and *arithmetic residuals*. A perfectly reversible, noiseless system can produce arbitrarily deep statistical memory when observed through a coarse lens.

Appendix A: Proof of Theorem 1 (Complete)

This appendix restates the existence theorem with full technical details.

A.1 Setup

Let $\beta \notin \mathbb{Q}$. For integer $m \geq 2$, define $x_{t+1} = (x_t + \beta) \bmod 1$, $\omega_t = \lfloor m x_t \rfloor$, $r_t = \{m x_t\}$, $A(m) = m\beta = k(m) + a(m)$. Then $\omega_{t+1} \equiv \omega_t + k(m) + c_t \pmod m$ where $c_t = 1\{r_t \geq 1\text{-}a(m)\}$.

A.2 Lemma (carry is non-Markov for small a)

For $a = a(m) \in (0, 1/3)$:

- Step 1: $c_t = 1 \Longrightarrow r_{t+1} \in [0,a) \Longrightarrow c_{t+1} = 0$. So $P(c_{t+1}=1|c_t=1) = 0$.
- Case A: $c_{t-1}=1, c_t=0 \Longrightarrow r_t \in [0,a)$. Condition $c_{t+1}=1$ requires $r_t \in [1\text{-}2a, 1\text{-}a)$, disjoint from $[0,a)$ for $a < 1/3$. So $P = 0$.
- Case B: $c_{t-1}=0, c_t=0 \Longrightarrow r_t \in [a, 1\text{-}a)$. Condition $c_{t+1}=1$ requires $r_t \in [1\text{-}2a, 1\text{-}a) \subset [a, 1\text{-}a)$. Length ratio $= a/(1\text{-}2a) > 0$.
- The two conditional probabilities differ, so $\{c_t\}$ is not first-order Markov. □

A.3 Existence

By Weyl equidistribution, $\{m\beta\}$ is dense in $[0,1)$. Hence infinitely many m have $a(m) \in (0, 1/3)$. Rationals in $(0,1/3)$ are countable; infinitely many m remain with $a(m)$ irrational.

A.4 Bijection

Given (ω_t, c_t): $\omega_{t+1} = \omega_t + k(m) + c_t \pmod m$ is unique. Given (ω_t, ω_{t+1}): $c_t = (\omega_{t+1} - \omega_t - k(m)) \bmod m \in \{0,1\}$ is unique. So $\{\omega_t\}$ is Markov iff $\{c_t\}$ is. Since $\{c_t\}$ is not Markov, $I(\omega_{t-1}; \omega_{t+1} \mid \omega_t) > 0$. □

Appendix B: Computational Methods

B.1 I_k Estimation

Conditional mutual information $I_k = I(\omega_{t-k}; \omega_{t+1} \mid \omega_{t-k+1..t})$ is estimated by exact counting using byte-key hashing (void-key method) to avoid integer overflow for large m and k. Sequence lengths T = 200,000–300,000 were used throughout.

B.2 Control Experiment Parameters

$\alpha = \pi^2$, $\gamma = \pi/3$, $m \in [20,300]$, $q \in \{3,5\}$, seeds $\in \{0,1,2\}$, T = 200,000, p = 0.25 for Bernoulli mode.

B.3 Markov Order Scan Parameters

$\beta = \pi^2/(2\pi)$ (corresponding to $\alpha = \pi^2$), $m \in [20,400]$, $K_{max} = 10$, $\varepsilon = 10^{-4}$ bits, T = 300,000.

Appendix C: Single-Scale Infinite Markov Order via Geometric Follower-Set Separation

This appendix proves the strongest result in the paper: for any fixed m with a(m) irrational, the induced process $\{\omega_t\}$ has infinite Markov order. The proof connects the carry sequence to Sturmian sequence theory via a geometric argument on circle partitions.

C.1 Rotation Coding and the Factor Partition

Fix $m \geq 2$ and write $a = a(m) = \{m\beta\} \in (0,1)$, assuming $a \notin \mathbb{Q}$. The residual rotation $r_{t+1} = (r_t + a) \bmod 1$ produces the binary carry coding:

$$c_t = 1\{r_t \in I_1\}, \quad I_1 = [1-a, 1), \quad I_0 = [0, 1-a)$$

Equivalently, for intercept $\rho = r_0$:

$$c_t(\rho) = 1\{\rho + ta \in I_1\} \pmod 1$$

For any word $w = w_0 \cdots w_{n-1} \in \{0,1\}^n$, define its cylinder set:

$$J(w) = \{\rho \in [0,1) : c_0(\rho)=w_0, \ldots, c_{n-1}(\rho)=w_{n-1}\}$$

Each constraint $c_t(\rho) = 1$ or 0 restricts ρ to a translate of I_1 or I_0 on the circle. Hence $J(w)$ is an intersection of half-open intervals — itself a half-open interval.

The only boundary points that appear are preimages of the partition boundary 1-a. For times $t = 0,\ldots,n-1$, the relevant boundary set is:

$$B_n = \{ (1-a) - ta \pmod 1 : t = 0,\ldots,n-1 \}$$

Since a is irrational, all n points in B_n are distinct. Sorting B_n on the circle partitions $[0,1)$ into exactly n+1 arcs. Each nonempty cylinder $J(w)$ is exactly one such arc.

> **Geometric takeaway:** Length-n factors are in bijection with the arcs of a circle partition by n rotated cut points. This is the standard geometric picture of Sturmian sequences.

C.2 The Unique Right-Special Interval and Nested Refinement

A length-n factor w is right-special if both w0 and w1 occur as length-(n+1) factors.

Geometrically: appending the next symbol $c_n(\rho)$ depends on whether ρ lies to the left or right of the next cut point:

$$b_n = (1-a) - na \pmod 1$$

Among all length-n cylinders, exactly one interval J_n contains b_n in its interior. That interval's factor u_n has two possible extensions, making it the unique right-special word. All other cylinders lie entirely on one side of b_n and have a unique extension.

Passing from length n to n+1 adds cut point b_n and splits the right-special cylinder:

$$J(u_n) = J(u_{n+1}) \cup K_n \quad \text{(disjoint)}$$

where $J(u_{n+1})$ is the unique right-special cylinder at length n+1. In particular:

$$J(u_{n+1}) \subsetneq J(u_n) \quad \text{for all } n$$

> **Geometric takeaway:** There is exactly one ambiguous cylinder at each length, and these form a strictly nested chain. This is the defining property of Sturmian sequences: factor complexity $p(n) = n+1$.

Numerical verification ($\beta = \pi^2$, $m = 6$, $T = 500{,}000$):

Length n	Distinct factors	Predicted p(n)=n+1	Right-special count
1	2	2	1
2	3	3	1
3	4	4	1
5	6	6	1
8	9	9	1
10	11	11	1
14	15	15	1

C.3 Infinitely Many Distinct Follower Sets → Non-Regular

For a word w, define its follower set:

```
F(w) = { v ∈ {0,1}* : wv ∈ L }
```

where L is the factor language of $\{c_t\}$.

> **Claim:** The follower sets $F(u_n)$ are all distinct for different n. Hence the collection $\{F(w)\}$ is infinite.

Proof of Claim

Pick any point $\rho_n \in K_n = J(u_n) \setminus J(u_{n+1})$ (nonempty because the inclusion is strict). Consider the tail v_n defined as future symbols starting at time n:

```
v_n := c_n(ρ_n) c_{n+1}(ρ_n) c_{n+2}(ρ_n) ...
```

Since $\rho_n \in J(u_n)$, the concatenation $u_n v_n^{(\ell)}$ occurs in L for any finite prefix length ℓ. Hence $v_n^{(\ell)} \in F(u_n)$.

But $\rho_n \notin J(u_{n+1})$, and $J(u_{n+1})$ is exactly the set of intercepts whose first n+1 symbols equal u_{n+1}. The refinement cut at b_n forces the next symbol after u_n to be fixed inside $J(u_{n+1})$, while ρ_n lies in the opposite half K_n. Therefore $v_n^{(\ell)} \notin F(u_{n+1})$.

Hence $F(u_n) \neq F(u_{n+1})$ for every n, so infinitely many distinct follower sets exist. □

Now invoke the Myhill-Nerode criterion: a language is regular if and only if it has finitely many distinct right-quotient (follower) sets. Since $\{c_t\}$ has infinitely many, its factor language L is not regular.

C.4 From Non-Regular to Infinite Markov Order

A sofic shift admits a finite labeled graph presentation; hence its factor language is regular. Since L is not regular, the induced subshift of $\{c_t\}$ is not sofic.

Any finite-order-L Markov process over a finite alphabet can be represented as a finite-state edge shift (take length-L contexts as states), hence is sofic. Therefore $\{c_t\}$ cannot be any finite-order Markov chain.

Measure-theoretic bridge. The argument above establishes that the carry shift is not sofic as a topological dynamical system. Since the rotation coding of $\{c_t\}$ is a Sturmian system, it is minimal and uniquely ergodic: there is exactly one shift-invariant Borel probability measure on the subshift. Therefore any finite-order Markov representation of the stochastic process would require a sofic topological model (via the unique ergodic measure), which is impossible. The non-regularity result thus transfers directly from the topological to the measure-theoretic (probabilistic) setting.

> **Conclusion:** $\{c_t\}$ has infinite Markov order at this fixed m, for any m with $a(m) \notin \mathbb{Q}$.

C.5 Transfer to $\{\omega_t\}$ via Recoverability

From the bijection established in Section 3.2: c_t is a finite-window function of ω (specifically $c_t = f(\omega_t, \omega_{t+1})$). If $\{\omega_t\}$ were L-step Markov for some finite L, then $c_t = f(\omega_t, \omega_{t+1})$ would also be finite-order Markov (order at most L+1). This contradicts the infinite-order result for $\{c_t\}$.

> **Theorem 3 (Infinite Markov order at fixed scale).** *For any irrational β and any m with $a(m) = \{m\beta\} \notin \mathbb{Q}$, the coarse-grained process $\{\omega_t\}$ has infinite Markov order. By Weyl equidistribution, such m constitute a set of full density in the positive integers.*

References

The following references provide background for the methods and results used in this paper.

Projection-Induced Memory and Mori-Zwanzig Formalism

[1] Nakajima, S. (1958). On quantum theory of transport phenomena. *Progress of Theoretical Physics*, 20(6), 948–959.

[2] Zwanzig, R. (1960). Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5), 1338–1341.

[3] Mori, H. (1965). Transport, collective motion, and Brownian motion. *Progress of Theoretical Physics*, 33(3), 423–455.

[4] Zwanzig, R. (2001). *Nonequilibrium Statistical Mechanics*. Oxford University Press.

[5] Chorin, A. J., Hald, O. H., & Kupferman, R. (2000). Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences*, 97(7), 2968–2973.

Symbolic Dynamics and Sturmian Sequences

[6] Morse, M., & Hedlund, G. A. (1940). Symbolic dynamics II. Sturmian trajectories. *American Journal of Mathematics*, 62(1), 1–42.

[7] Lind, D., & Marcus, B. (1995). *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press.

[8] Weiss, B. (1973). Subshifts of finite type and sofic systems. *Monatshefte für Mathematik*, 77(5), 462–474.

[9] Lothaire, M. (2002). *Algebraic Combinatorics on Words*. Cambridge University Press. (Chapter 2: Sturmian Words).

Equidistribution and Number Theory

[10] Weyl, H. (1916). Über die Gleichverteilung von Zahlen mod. Eins. *Mathematische Annalen*, 77(3), 313–352.

Automata Theory and Formal Languages

[11] Nerode, A. (1958). Linear automaton transformations. *Proceedings of the American Mathematical Society*, 9(4), 541–544.

Information Theory

[12] Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley-Interscience.

Ergodic Theory

[13] Walters, P. (1982). *An Introduction to Ergodic Theory*. Springer Graduate Texts in Mathematics, Vol. 79.