Unpredictability from Parity: Understanding Linear Correctors for Markovian Sources

Maciej Skorski* Johannes Mittmann†
*University of Warsaw, University of Luxembourg †Bundesamt für Sicherheit in der Informationstechnik (BSI)

***Abstract*—Linear correctors are widely applied to reduce the bias of bit streams via parity (XOR), well beyond their security guarantees: these are established mostly for independent coins, which physical sources with memory (e.g. phase noise) violate.**

We close this gap by developing a *fully general, exact formula* for the parity bias of dependent bits from a (hidden, possibly non-homogeneous) Markov chain: in this formula, the bias generating function factorizes as a product of "sign-twisted" transfer kernels.

Under next-bit unpredictability, i.e., conditional bias at most β, the decay is geometric in the number of bits used, with per-bit rate of order $\beta^{1/(m+1)}$ for m-bit memory, between explicit constants: 1 from below and $(m+1)^{1/(m+1)} \le 3^{1/3} \approx 1.44$ from above. The gap between constants closes for first-order chains, and closes with growing m. We note that *bounded memory is necessary*: under bounded unpredictability but the full history, XOR-ing does not reduce the bias. As an application note: the exact bias may oscillate around its geometric decay and should be certified at the exact finite operating length for deployment.

Compared with the best previous analysis (Zhou and Bruck, ISIT 2011), we fix a flaw in their key lemma, tighten the bound exponentially, and cover higher-order chains.

***Index Terms*—randomness extraction, bias correction, Markov chains, XOR correctors, parity, linear extractors**

I. Introduction

Linear correctors use the parity (XOR) of bits to reduce their bias. Proposed by Dichtl [1] as a lightweight remedy for biased physical randomness, they remain a standard post-processing step in hardware random number generators, for example in FPGA TRNGs built on XORed ring oscillators [2], or, in the form of Toeplitz hashing, high-throughput quantum RNGs [3]; TRNG standards treat post-processing as a component to be evaluated [4], [5].

Yet rigorous guarantees, needed for security, are almost entirely limited to independent bits: the classical Piling-up Lemma [6] gives the geometric bias-decay rate for independent coins, and the follow-up literature largely keeps that assumption [7]–[10]. Raw bits from physical sources are not independent, however: in popular entropy sources, such as thermal noise in free-running ring oscillators, the sampled phase performs a random walk [11], and oscillator noise is often exponentially correlated (Ornstein–Uhlenbeck) [12], so the output is naturally Markovian.

This gap between what theory guarantees and what practice assumes motivates a problem of both practical interest (TRNGs) and foundational interest (parity as the bias of a sum of dependent bits):

Research question. *How effective is parity in removing bias for Markov sources, hidden or directly observed?*

We close this research gap by computing the exact parity bias under (hidden) Markov dependencies, and the worst-case decay rates when the next-bit bias is bounded by β.

A. Our Contributions

Throughout, we study bits $X_i = f(\theta_i)$ obtained from a first-order Markov chain (θ_i) on a countable state space. This covers a chain observed directly ($\theta = X$, $f = \mathrm{id}$), a chain of order m (the state is the context of m previous bits, f its last bit), and hidden chains with post-processing f. Of interest are sum parity $S_n = X_1 \oplus \cdots \oplus X_n$ and particularly their bias $\mathsf{bias}(S_n) = \mathbb{E}(-1)^{S_n}$. The source quality is quantified by its *next-bit bias* β, $|\mathbb{E}[(-1)^{X_i} \mid \theta_{\le i-1}]| \le \beta$, where $\beta = 0$ corresponds to a fair coin and $\beta = 1$ is deterministic; all rates below are stated in terms of β. We obtain the following results:

1) **Generating function for general Markov sums.** For first-order sources (any state space, possibly non-homogeneous), the probability generating function factorizes into *twisted kernels* used as transfer matrices; order-m chains are first-order on their contexts (Theorem III.1; Corollaries A.1 and A.2 in the appendix):

$$\mathbb{E}\, t^{X_1+\cdots+X_n} = \pi \prod_{i=1}^{n} \mathbf{P}^{(i)} \mathbf{Z}_t\, \mathbf{1}, \quad \mathbf{Z}_t = \mathrm{diag}(t^f).$$

   Here π is the initial-state row vector, $\mathbf{P}^{(i)}$ the chain's transition matrix at step i, and $\mathbf{1}$ the all-ones column vector. For $t = -1$ and homogeneous chains the parity bias is the compact matrix power $\mathsf{bias}(S_n) = \pi \mathbf{K}^n \mathbf{1}$ with $\mathbf{K} = \mathbf{P}\mathbf{Z}_{-1}$ (Corollary III.6). For hidden sources, a two-sided next-bit bias β_2 on the middle bit given both neighboring states gives $|\mathsf{bias}(\bigoplus_j X_{i_j})| \le \beta_2^{\lfloor (q-1)/2 \rfloor}$ for any q selected positions (Corollary III.2); this corrects the exponent in Zhou and Bruck's [13] Lemma 7 (Proposition III.5 and corollary III.4).
2) **Exact dynamic for binary sources.** For a binary first-order Markov chain the twisted kernel is 2×2, and the Cayley-Hamilton relation $\mathbf{K}^2 = \tau \mathbf{K} + \lambda \mathbf{I}$ (with $\tau = \mathrm{tr}(\mathbf{K})$, $\lambda = -\det(\mathbf{K})$) gives a Fibonacci-type recursion with an explicit formula (Theorems III.8 and III.10):

$$\mathsf{bias}(S_n) = \lambda F_{n-1} + F_n\ \mathsf{bias}(S_1), \quad F_n = \tau F_{n-1} + \lambda F_{n-2}.$$

The bias decays at the geometric rate $\rho(\mathbf{K})^n$, up to a factor of $n+1$. If the next-bit bias is $\le \beta$, then $|\,\mathsf{bias}(S_n)| \le \beta^{\lceil n/2 \rceil}$, attained for every n (Theorem III.11).

3) **Near-optimal per-bit rate for bounded memory.** For order-m chains with next-bit bias at most β, the spectral radius of the twisted kernel, which governs the per-bit decay of the parity bias, obeys (Proposition III.16)

$$\rho(\mathbf{K}) \le \min\left\{1, \big((m+1)\beta\big)^{1/(m+1)}\right\}.$$

The parity bias of a noisy linear-feedback-shift-register (LFSR [14]) decays at exactly $\beta^{1/(m+1)}$ per bit (Proposition III.14), so no bound in terms of β can beat this rate; ours is within a factor $(m+1)^{1/(m+1)} \le 3^{1/3} \approx 1.44$ of it for every m, hence of the best possible rate. Moreover, $\rho(\mathbf{K})^{m+1} \le \beta(1 + O(\beta^{1/(m+1)}))$ (Proposition III.16), so the rate is asymptotically exact.

4) **Memory is necessary:** we show that when next-bit unpredictability is imposed on the full history, XOR-ing does not reduce the bias at all (Section III-D).

B. Related Work

Independent inputs. XOR-ing for bias reduction goes back to Davies [15]; linear correctors were formalized by Dichtl [1] and analyzed for independent bits by Lacharme [7] and Zhou and Bruck [8], and, most recently, by Skórski, Soto and Günlü [10], who give exact formulas and near-tight distance-to-uniformity bounds via code weight enumerators. All these results reduce to the parity biases $\mathsf{bias}(\sum_{i\in I} X_i)$ of the input, which factor for independent bits by the Piling-up Lemma [6]; for dependent inputs, we compute them for Markov sources. **Dependent inputs.** Zhou and Bruck [8], [13] give the best previous bound for Markov sources: for hidden Markov sources [16], including order-one chains, their Lemma 9 (Lemma 7 in the extended version) claims $|\,\mathsf{bias}\,| \le e^{(q-1)/2}$ for q bits, with $e = 2\beta/(1+\beta^2)$. We improve this in exponent, tightness, and scope: (i) *exponent*: theirs is false in general (Proposition III.5), and we prove the correct one, $\lfloor (q-1)/2 \rfloor$ (Corollaries III.2 and III.4); (ii) *tightness*: for binary chains the base improves to β and the exponent to $\lceil n/2 \rceil$ on n consecutive bits, a gain of about $2^{(n-1)/2}/\beta$, and the bound is sharp (Remark III.3 and theorem III.11); (iii) *scope*: for order $m \ge 2$ their two-sided hypothesis degenerates ($\theta_{j+1} \ni X_j$ forces $e = 1$), whereas the one-sided β suffices for our rate $\beta^{1/(m+1)}$ (Propositions III.14 and III.16).

Simion [17] gives an exact 4×4 matrix expression for the parity of a binary homogeneous chain, which is a special case of our recursion; it yields no decay rates and no higher-order or non-homogeneous chains, and two of his closed forms are wrong (cases 2 and 4 of [17, Sec. 2]; corrected via Theorem III.8). Extraction from Markov sources by non-linear, variable-length procedures (von Neumann, Elias, Peres, Blum) is classical [18]–[24]; we analyze fixed-rate linear correctors, preferred in hardware; Santha–Vazirani sources admit no deterministic extractor [25], [26].

Markovian counting processes. Edwards [27] derives the generating function of the number of ones in n bits of a *binary, homogeneous, first-order* chain. Its value at $z = -1$ is the parity bias only in this special case, which is our $m = 1$ binary setting, and parity is not studied there. Our Theorem III.1 covers arbitrary state spaces, hidden and non-homogeneous chains, and (via the state of m bits) any order m.

II. Preliminaries

The *bias* of a binary random variable Z is $\mathsf{bias}(Z) = \mathbb{P}(Z=0) - \mathbb{P}(Z=1) = \mathbb{E}[(-1)^Z]$; it determines the law of Z. For independent X_j, parity multiplies biases (Piling-up Lemma [6]): $\mathsf{bias}(\bigoplus_{j\in J} X_j) = \prod_{j\in J} \mathsf{bias}(X_j)$. With $S_0 = 0$, we also use $S_n^+ = S_n \oplus \mathrm{wt}(c_0)$, the parity that includes the initial context ($= X_0 \oplus \cdots \oplus X_n$ for $m = 1$).

For a source $(X_n)_{n\ge 1}$ with (possibly hidden) state θ, the (conditional) *next-bit bias* is

$$\beta := \sup_{n\ge 1,\, \theta_{\le n-1}} \big|\mathbb{E}\big[(-1)^{X_n} \mid \theta_{\le n-1}\big]\big|,$$

that is, $\mathbb{P}(X_n = x \mid \theta_{\le n-1}) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$; $\beta = 0$ is a fair source and $\beta = 1$ a deterministic one. For observed sources $\theta = X$, and conditioning on the whole history gives Santha–Vazirani sources [25].

A binary first-order chain has transition matrix $\mathbf{P} = \begin{pmatrix} \delta_1 & 1-\delta_1 \\ 1-\delta_2 & \delta_2 \end{pmatrix}$ with *holding probabilities* $\delta_1, \delta_2 \in [0,1]$.

III. Main Results

We go from general to specific: the generating-function identity and its order-m form, the solved one-bit case, rate bounds for bounded memory, and the barrier for unbounded memory. Table I summarizes the worst-case per-bit decay rate of the parity bias under next-bit unpredictability β. Proofs are in the appendix; the full paper is IACR ePrint 2026/2417 [28], simulations are in the repository [29].

Intuition. Once we express the parity bias as a product of twisted kernels (one-step transitions signed by $(-1)^{\text{emitted bit}}$), it remains to understand how fast it contracts. For independent bits, every XOR multiplies the bias by that of the new bit. With memory, a bit can be related to the m bits before it, so that XOR does not remove the bias of β. Hence one factor β is gained per block of $m+1$ bits, reflected in the rate $\beta^{1/(m+1)}$; with unbounded memory, where the last bit can copy the parity of all earlier ones, the bias stays β for every n.

A. Generating Function of Markovian Sums

Let $(\theta_i)_{i\ge 0}$ be a first-order Markov chain on an arbitrary state space Θ with (possibly time-dependent) kernels $\mathbf{P}^{(i)}$ and $\theta_0 \sim \pi$, and let $X_i = f(\theta_i) \in \{0,1\}$. For $t \in \mathbb{C}$, let $\mathbf{Z}_t = \mathrm{diag}(t^f)$ have diagonal entries $t^{\text{emitted bit}}$. The *twisted kernel* $\mathbf{K}_t^{(i)} = \mathbf{P}^{(i)}\mathbf{Z}_t$ is one step of the chain, weighted by

TABLE I
WORST-CASE PARITY-BIAS DECAY RATES; β, β_2 BOUND THE NEXT-BIT AND TWO-SIDED BIASES; THE DECAY IS NOT MONOTONE (FIGURE 1).

Memory, assumption	Ref.	Rate r [finite n]
none (i.i.d.), β	[6]	$r = \beta$ [β^n]
order 1, binary, β	III.11	$r = \sqrt{\beta}$ [$\beta^{\lceil n/2 \rceil}$, sharp]
order m, β	III.14 and III.16	$\sqrt[m+1]{\beta} \le r \le \sqrt[m+1]{(m+1)\beta}$
hidden, two-sided β_2	III.2	$r \le \sqrt{\beta_2}$ [$\beta_2^{\lfloor (n-1)/2 \rfloor}$]
unbounded, full history β	III.18	$r = 1$ [bias β for every n]

$t^{\text{emitted bit}}$ of the state reached. At $t = -1$ (entries $(-1)^{\text{emitted bit}}$) we write $\mathbf{Z} = \mathbf{Z}_{-1}$ and $\mathbf{K}^{(i)} = \mathbf{P}^{(i)}\mathbf{Z}$, i.e. $\mathbf{K}^{(i)}(\theta, d\theta') = \mathbf{P}^{(i)}(\theta, d\theta')\,\sigma(\theta')$ with $\sigma = (-1)^f$.

Theorem III.1 (Generating function of hidden Markov sources)**.** *With the notation as above,*

$$\mathbb{E}\, t^{X_1+\cdots+X_n} = \pi \mathbf{K}_t^{(1)} \cdots \mathbf{K}_t^{(n)} \mathbf{1}.$$

In particular $\mathsf{bias}(S_n) = \pi \mathbf{K}^{(1)} \cdots \mathbf{K}^{(n)} \mathbf{1}$.

The identity follows by induction on n: conditioning on θ_1 and using the Markov property gives $\mathbb{E}[t^{X_1+\cdots+X_n} \mid \theta_0] = \mathbf{K}_t^{(1)}\, \mathbb{E}[t^{X_2+\cdots+X_n} \mid \theta_1]$ (Appendix A).

Corollary III.2 (Bounded two-sided noise)**.** *In the setting above, suppose that for some $\beta_2 \in [0,1]$ and all times $i < j < k$ (including the last time),*

$$\left|\mathbb{E}\big[(-1)^{X_j} \mid \theta_i, \theta_k\big]\right| \le \beta_2. \tag{1}$$

Then $\|\mathbf{K}^{(j)}\mathbf{K}^{(j+1)}\|_\infty \le \beta_2$ for all $j \ge 2$ (for $j = 1$ if θ_0 is included), and for any positions $i_1 < \cdots < i_q$,

$$\left|\,\mathsf{bias}\Big(\bigoplus_{j=1}^{q} X_{i_j}\Big)\right| \le \beta_2^{\lfloor (q-1)/2 \rfloor}.$$

Unlike the next-bit bias β, the two-sided bias β_2 also conditions on a future state; for hidden chains no bound on β alone suffices (Section III-D).

The proof idea is as follows: two consecutive twisted kernels average a middle bit given both neighbors, so they contract by β_2, and every second selected position costs one factor β_2.

Remark III.3 (Relation to Zhou and Bruck [13]). Their model is the hidden Markov source $X_i = f(\theta_i)$, with the noise hypothesis (1) (parameter e) imposed for all $i < j < k \le n$ in Lemma 9 of the ISIT paper [8] (and $k < n$ in Lemma 7 of the extended version [13]); it is the hypothesis of Corollary III.2. Under this hypothesis, (1) charges every pair of neighbors at its worst case, giving $\beta_2^{\lfloor (q-1)/2 \rfloor}$ (it conditions on both neighbors, so the last selected bit of an even-size set is not covered), whereas they claim the stronger $\beta_2^{(q-1)/2}$, which Proposition III.5 shows is false in general. For a first-order binary chain ($\theta = X$, $\mathbb{P}(X_i \mid X_{i-1}) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$) their assumption holds with $\beta_2 = 2\beta/(1+\beta^2)$, whereas the exact value is $\|\mathbf{K}^2\|_\infty = \beta$, attained for $\delta_1 = \delta_2$ (Cayley–Hamilton, $\mathbf{K}^2 = \tau\mathbf{K} + \lambda\mathbf{I}$). For n consecutive bits their claimed bound exceeds ours by the factor $(2/(1+\beta^2))^{(n-1)/2}\beta^{-1}$ (n odd; $\beta^{-1/2}$ in place of β^{-1} for n even), about $2^{(n-1)/2}/\beta$ for small β. Our bound is attained for every n (Theorem III.11), and for observed chains ours needs only the causal next-bit bias β, whereas their hypothesis also conditions on future states.

Corollary III.4 (Zhou and Bruck's Lemma 7, Corrected)**.** *Under the hypothesis of [13] that $\left|\mathbb{P}[x_{i_2} = 1 \mid \theta_{i_1}, \theta_{i_3}] - \frac{1}{2}\right| \le \frac{e}{2}$ for all $i_1 < i_2 < i_3$, Corollary III.2 with $\beta_2 = e$ gives,*

$$\left|\mathbb{P}\Big[\bigoplus_{\ell=1}^{q} x_{i_\ell} = 1\Big] - \tfrac{1}{2}\right| \le \tfrac{1}{2} e^{\lfloor (q-1)/2 \rfloor},$$

for any selected positions $i_1 < \cdots < i_q$.

Proposition III.5 (Lemma 9 of [8] (Lemma 7 of [13]) is false as stated)**.** *Let A, C be independent fair bits and let $(x_1, x_2, x_3) = (A, C, A)$ with hidden states $\theta_1 = A$, $\theta_2 = (A, C)$, $\theta_3 = A$. This hidden Markov source satisfies the noise hypothesis of [13] with $e = 0$, yet $x_1 \oplus x_3 \equiv 0$, so $|\mathbb{P}[x_1 \oplus x_3 = 1] - \frac{1}{2}| = \frac{1}{2} > \frac{1}{2}e^{1/2}$ for every $e < 1$, contradicting their lemma for $q = 2$. The same holds for every length $n \ge 3$ and under either index convention (Appendix D); for odd q the two exponents coincide, so the correction concerns even q.*

We now specialize to homogeneous order-m binary chains, the setting for the rest of the paper.

Any order-m chain is first-order on its length m contexts $s = (s_1, \ldots, s_m)$ with $f(s) = s_m$ (the last symbol), so Theorem III.1 applies. Its kernel is $\mathbf{P}(s, s') = \mathbb{P}(x \mid s)$ if $s' = (s_2, \ldots, s_m, x)$ and 0 otherwise. With $\mathbf{Z}_t = \mathrm{diag}(t^{s_m})$ and $\mathbf{Z} = \mathbf{Z}_{-1}$, the twisted kernel is $\mathbf{K} = \mathbf{P}\mathbf{Z}$, i.e. $\mathbf{K}(s, s') = (-1)^x \mathbf{P}(s, s') = \frac{(-1)^x + b_s}{2}$ in terms of the next-bit bias $b_s = \mathbb{E}[(-1)^X \mid s]$.

Corollary III.6 (Homogeneous order-m chains)**.** *For a homogeneous order-m chain with initial context law π and $N_n = \sum_{i=1}^n X_i$ (so $S_n = N_n \bmod 2$),*

$$\mathbb{E}\, t^{N_n} = \pi\, (\mathbf{P}\mathbf{Z}_t)^n \mathbf{1}.$$

At $t = -1$,

$$\mathsf{bias}(S_n) = \pi\, \mathbf{K}^n \mathbf{1},$$
$$\mathsf{bias}(S_n \mid c_0) = (\mathbf{K}^n \mathbf{1})_{c_0},$$

and, with $\mathbf{D} = \mathrm{diag}((-1)^{\mathrm{wt}(s)})$,

$$\mathsf{bias}(S_n^+) = \pi\, \mathbf{D}\, \mathbf{K}^n \mathbf{1}.$$

B. One-Bit Memory: Exact and Sharp Results

We turn to the exactly solvable case of a single bit of memory ($m = 1$), where Theorem III.1 simplifies to a closed-form recursion and a sharp rate.

Corollary III.7 (One-bit memory)**.** *For a homogeneous chain on $\{0,1\}$ with $\mathbf{P}$ in holding-probability form,*

$$\mathbf{K} = \mathbf{P}\mathbf{Z} = \begin{pmatrix} \delta_1 & -(1-\delta_1) \\ 1-\delta_2 & -\delta_2 \end{pmatrix},$$

$$\mathsf{bias}(S_n \mid X_0 = x) = (\mathbf{K}^n \mathbf{1})_x.$$

By the Markov property, the same formula gives the bias of the parity of any block of n consecutive bits, given the bit just before it (Corollary III.13).

This is Corollary III.6 for $m = 1$; it is proved in Appendix G in two ways, by induction and by linear algebra over the augmented (X_n, S_n) chain.

We now use a more convenient *balance-correlation* parametrization of $\mathbf{P}$. With next-bit biases $b_x = \mathbb{E}[(-1)^{X_{n+1}} \mid X_n = x]$, i.e. $b_0 = 2\delta_1 - 1, b_1 = 1 - 2\delta_2$, we put

$$\begin{aligned} \tau &= \operatorname{tr} \mathbf{K} = \delta_1 - \delta_2 = \tfrac{b_0+b_1}{2}, \\ \lambda &= -\det \mathbf{K} = \delta_1 + \delta_2 - 1 = \tfrac{b_0-b_1}{2}, \end{aligned} \tag{2}$$

so that $b_x = \tau + \lambda(-1)^x$: τ is the average next-bit bias, and λ, the part that depends on X_n, is the lag-1 autocorrelation under the stationary law. The next-bit bias of the chain is $\beta = \max_x |b_x| = |\tau| + |\lambda|$, i.e. at most β iff $\delta_1, \delta_2 \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$.

Theorem III.8 (Recursion and Binet representation)**.** *Let $S_n = \bigoplus_{i=1}^n X_i$ (with $S_0 := 0$, so $\mathsf{bias}(S_0) = 1$). Then*

$$\mathsf{bias}(S_n) = \tau\, \mathsf{bias}(S_{n-1}) + \lambda\, \mathsf{bias}(S_{n-2}), \qquad n \geq 2, \tag{3}$$

$$\mathsf{bias}(S_n) = \lambda F_{n-1} + F_n\, \mathsf{bias}(S_1), \qquad n \geq 0, \tag{4}$$

where $F_0 = 0$, $F_1 = 1$, $F_n = \tau F_{n-1} + \lambda F_{n-2}$, with $\lambda F_{-1} := 1$, is the generalized Fibonacci sequence [30] for τ, λ.

The proof uses Cayley–Hamilton's formula [31] $\mathbf{K}^2 = \tau\mathbf{K} + \lambda\mathbf{I}$, replacing matrix powers with a two-term recursion.

Corollary III.9 (Two Extreme Chains)**.** *Two special cases highlight the distinct roles of τ and λ: the correlation λ alone gives the square-root rate $|\lambda|^{1/2}$ per bit, while the balance τ alone gives the independent rate $|\tau|$.*

- Balanced chain, $\tau = 0$. *Here $F_n = \lambda F_{n-2}$, so $F_{2k} = 0$ and $F_{2k+1} = \lambda^k$. The above Binet formula then gives*

$$\mathsf{bias}(S_n) = \begin{cases} \lambda^{n/2}\, \mathsf{bias}(S_0), & n \text{ even}, \\ \lambda^{(n-1)/2}\, \mathsf{bias}(S_1), & n \text{ odd}. \end{cases}$$

  So $|\mathsf{bias}(S_n)| = |\lambda|^{n/2}$ for even n; for a stationary start with $\lambda < 1$, $\mathsf{bias}(S_1) = 0$ and the odd-step bias vanishes identically.
- Memoryless chain, $\lambda = 0$. *The recursion collapses to $\mathsf{bias}(S_n) = \tau\, \mathsf{bias}(S_{n-1})$, giving $\mathsf{bias}(S_n) = \tau^n\, \mathsf{bias}(S_0)$ — exactly matching the piling-up lemma.*

The eigenvalues of $\mathbf{K}$ give the exact asymptotics.

Theorem III.10 (Spectral formula)**.** *Let η_1, η_2 be the (possibly complex) eigenvalues of $\mathbf{K}$ (the roots of $t^2 - \tau t - \lambda = 0$), ordered so that $\rho := |\eta_1| \geq |\eta_2|$, and set $h := \eta_2/\eta_1$ ($h := 0$ if $\rho = 0$). Let $\Sigma_n = \sum_{i=0}^n h^i$. Then for $S_n = \bigoplus_{i=1}^n X_i$ and every $n \geq 1$,*

$$\max_x |\,\mathsf{bias}(S_n \mid X_0 = x)| = \rho^n |\Sigma_n| + \rho^{n+1} |\Sigma_n - 1|$$

and in particular, denoting $E_n = \rho^n(\rho + |\Sigma_n|)$,

$$\frac{E_n}{1+\rho} \leq \max_x |\,\mathsf{bias}(S_n \mid X_0 = x)| \leq (1+\rho) E_n.$$

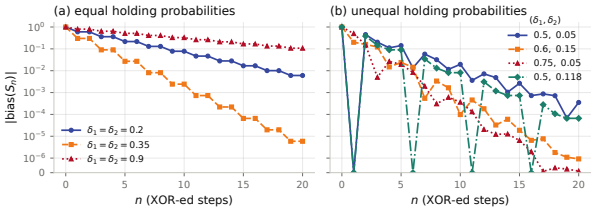

Fig. 1. Decay of $|\,\mathsf{bias}(S_n)|$, for chains with equal (left) and unequal (right) holding probabilities (δ_1, δ_2), starting at $X_0 = 0$;

Figure 2 shows how the rate ρ depends on the parameters, and Figure 3 the *leading factor* $\mathsf{bias}(S_n)/\rho^n$, which has three regimes. Its magnitude *oscillates* in $[0, (1+\rho)/|\sin(\theta/2)|]$ if $|h| = 1$, $\theta = \arg h \neq 0$ (the fan of sign changes in Figure 3; near-zeros and rebounds in Figure 1), grows up to $(1+\rho)n+1$ if the eigenvalues coincide ($h = 1$, i.e. $\tau^2 + 4\lambda = 0$; the ridge in Figure 3), and converges if $|h| < 1$. Unlike for independent bits with monotone decay, a guarantee at one length can thus mislead: it should be ensured at every length of interest.

Theorem III.11 (Optimal Bounds under Bounded Bias)**.** *For every binary Markov chain, every initial distribution of X_0,*

$$|\,\mathsf{bias}(S_n)| \leq \max_\pi |\,\mathsf{bias}(S_n)| = \|\mathbf{K}^n \mathbf{1}\|_\infty \leq \beta^{\lceil n/2 \rceil}. \tag{5}$$

The bound is attained for every n by $\delta_1 = \delta_2 = \frac{1\pm\beta}{2}$ and deterministic X_0.

Remark III.12. By Cayley–Hamilton $\|\mathbf{K}^2\|_\infty \leq |\tau| + |\lambda| = \beta$ while $\|\mathbf{K}\|_\infty = 1$, so every two XOR-ed bits cost one factor β. Independent bits of bias at most β give β^n; memory one costs a square root in the per-bit rate, $\beta^{\lceil n/2 \rceil}$; e.g. $\beta = 0.1$, $n = 8$: 10^{-8} becomes 10^{-4}.

A TRNG post-processor [5] may XOR *blocks* of n raw bits; as our bound holds for every initial distribution of X_0, it transfers to each block given the past, so the output is again next-bit unpredictable, with exponentially smaller bias $\beta^{\lceil n/2 \rceil}$.

Corollary III.13 (Block XOR-ing of streams)**.** *Let (X_i) be a binary Markov chain with next-bit bias at most β, and let $Y_b = \bigoplus_{i=(b-1)n+1}^{bn} X_i$ be the parities of consecutive non-overlapping blocks. Then for all b,*

$$\left|\mathbb{E}\left[(-1)^{Y_b} \mid Y_{<b}\right]\right| \leq \beta^{\lceil n/2 \rceil},$$

i.e. $(Y_b)_b$ is a Santha–Vazirani source with bias $\beta^{\lceil n/2 \rceil}$.

C. General Order-m Chains: Rate Bounds

For general memory m, Corollary III.6 no longer reduces to a simple closed form; we instead provide lower and upper bounds on the per-bit rate via the spectral radius of $\mathbf{K}$.

Proposition III.14 (Lower bound)**.** *Let $X_i = (\bigoplus_{j=i-m}^{i-1} X_j) \oplus \varepsilon_i$ for $i \geq 1$ with independent ε_i of bias β (a noisy LFSR [14] with all-ones taps) and initial context $X_{-m+1} = \cdots = X_0 = 0$. Then $\mathsf{bias}(S_n) = \beta^{\lceil n/(m+1) \rceil}$.*

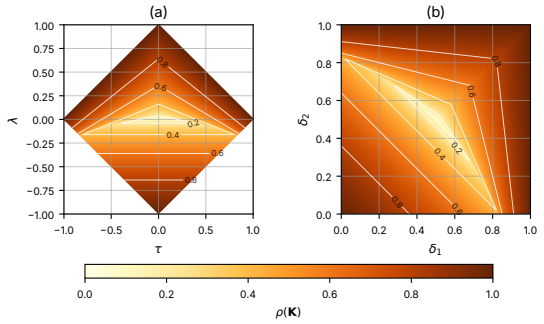

Fig. 2. Spectral radius $\rho(\mathbf{K})$ in terms of the two parametrizations: balance–correlation (τ, λ) (a) and holding probabilities (δ_1, δ_2) (b).

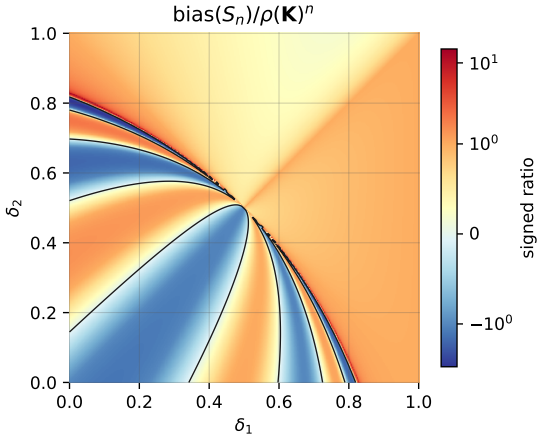

Fig. 3. Leading constant $C_n = \mathsf{bias}(S_n)/\rho(\mathbf{K})^n$ (signed, $n = 10, X_0 = 0$).

In this noisy LFSR, XORing $m+1$ consecutive bits leaves one independent noise bit, so each such group reduces the parity bias by a factor β. Unlike in correlation attacks [32], the noise enters the feedback, not the observed output.

Proposition III.15 (Elementary upper bound)**.** *For every order-m chain and every $n \geq 0$, $|\,\mathsf{bias}(S_n)| \leq \gamma^{\lfloor n/(m+1)\rfloor}$ with $\gamma = (1+\beta)^{m+1} - 1 \approx (m+1)\beta$, and the same bound holds for S_n^+.*

Proposition III.16 (Spectral radius via nilpotence)**.** *Let $L = m+1$. For every order-m chain, every eigenvalue η of $\mathbf{K}$:*

$$|\eta|^L \leq \beta\big(1 + |\eta| + \cdots + |\eta|^{L-1}\big).$$

Hence $\rho(\mathbf{K}) \leq r_*$, the unique nonnegative root of $r^L = \beta(1 + r + \cdots + r^{L-1})$ (for $\beta = 0$, $\mathbf{K}$ is nilpotent and $\rho(\mathbf{K}) = 0$), and

$$\rho(\mathbf{K}) \leq \min\{1, (L\beta)^{1/L}\} \leq 3^{1/3}\beta^{1/L} \leq 1.443\,\beta^{1/L}.$$

If $r_* < 1$, then moreover $\rho(\mathbf{K})^L \leq \beta/(1-r_*)$, so $\rho(\mathbf{K})^L \leq \beta(1 + O(\beta^{1/L}))$ as $\beta \to 0$.

By Proposition III.14 the noisy LFSR decays at exactly $\beta^{1/L}$ per bit, so the optimal per-bit rate lies in $[\beta^{1/L}, (L\beta)^{1/L}]$, within a factor $L^{1/L} \leq 3^{1/3} \approx 1.44$ for all m and β; the refined bound $\rho(\mathbf{K})^L \leq \beta(1 + O(\beta^{1/L}))$ of Proposition III.16 makes it asymptotically exact as $\beta \to 0$. The proof (Appendix O) replaces the binomial expansion of Proposition III.15 by the resolvent of the nilpotent part of $\mathbf{K}$ [33], the analogue of Cayley–Hamilton for $m = 1$.

Conjecture III.17. *For every $m \geq 1$ there is $\beta_m > 0$ such that for all $\beta \leq \beta_m$ and all order-m chains with $|b_s| \leq \beta$, $\rho(\mathbf{K}) \leq \beta^{1/(m+1)}$; equivalently the sharp per-bit rate is $\beta^{1/(m+1)}$, attained by the noisy LFSR of Proposition III.14.*

Numerics over all sign patterns $b_s = \pm\beta$ give $\sup \rho(\mathbf{K})^{m+1}/\beta = 1.000$ for $m = 2, 3$ (Appendix R).

D. Finite Memory is Necessary: Santha–Vazirani Barrier

Consider a class $\mathrm{SV}(\beta)$ of *Santha–Vazirani sources* [25] with bias β if $|\mathbb{E}[(-1)^{X_n} \mid X_{<n}]| \leq \beta$ almost surely for all n: the conditioning is on the whole history, as opposed to a bounded context. Parity filters then fail.

Theorem III.18 (Parity filters fail on Santha–Vazirani sources)**.** *For any finite non-empty set of indices A and $j = \max A$,*

$$\max_{(X_k)\in \mathrm{SV}(\beta)} \Big|\,\mathsf{bias}\Big(\bigoplus_{i\in A} X_i\Big)\Big| = \beta,$$

independently of $|A|$ and of the positions in A. The maximum is attained by a source that is fair at every position except j. In particular $\sup |\,\mathsf{bias}(S_n)| = \beta$ for every $n \geq 1$.

Corollary III.19 (Linear correctors do not help)**.** *Let $G \in \mathbb{F}_2^{k\times n}$ have a nonzero row and $Y = GX$, and let $\Delta_{\mathsf{SD}}(Y) = \frac{1}{2}\sum_y |\mathbb{P}(Y = y) - 2^{-k}|$ be its statistical distance from uniform. For every $\beta \in (0,1)$ there is a Santha–Vazirani source with bias β and $\Delta_{\mathsf{SD}}(Y) \geq \beta/2$ and $\max_y \mathbb{P}(Y = y) \geq 2^{-k}(1+\beta)$, i.e. $H_\infty(Y) \leq k - \log_2(1+\beta)$. Hence Y is no better than one raw bit, for every n and G.*

The extremal source makes X_j a noisy copy of the parity P of the earlier bits of A, $\mathbb{P}(X_j = P) = \frac{1+\beta}{2}$, so $P \oplus X_j$ has bias β; this needs memory spanning A; for order-m chains gains come only from disjoint blocks of $m+1$ bits, one factor β per block; with unbounded memory there is none.

IV. Conclusion

We gave a *fully general, exact formula* for the parity bias of Markov sources via *twisted kernels*, $\mathsf{bias}(S_n) = \pi\mathbf{K}^{(1)}\cdots\mathbf{K}^{(n)}\mathbf{1}$, and *near-optimal decay rates* under next-bit unpredictability: the *sharp bound* $\beta^{\lceil n/2\rceil}$ for first-order binary chains, whose decay is *non-monotone*, and a rate *within a factor* $3^{1/3} \approx 1.44$ of the best possible one for order-m chains. We *correct, extend and tighten* Zhou and Bruck's result by about $2^{\lfloor (n-1)/2\rfloor}/\beta$. Finally, we demonstrate that *bounded memory is necessary* for correcting with XOR.

Future work: total variation distance for $Y = GX$ via the XOR lemma (our formula provides all parity biases), and the sharp rate for higher orders (Conjecture III.17).

References

[1] Markus Dichtl, "Bad and good ways of post-processing biased physical random numbers," in *Fast Software Encryption, 14th International Workshop, FSE 2007, Luxembourg, Luxembourg, March 26–28, 2007, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 4593. Berlin, Heidelberg: Springer, 2007, pp. 137–152. [Online]. Available: https://doi.org/10.1007/978-3-540-74619-5_9

[2] Cristian Klein, Octavian Cret, and Alin Suciu, "Design and implementation of a high quality and high throughput TRNG in FPGA," 2009. [Online]. Available: https://arxiv.org/abs/0906.4762

[3] Xiaomin Guo, Fading Lin, Jiehong Lin, Zhijie Song, Yue Luo, Qiqi Wang, and Yanqiang Guo, "Parallel and real-time post-processing for quantum random number generators," 2024. [Online]. Available: https://arxiv.org/abs/2403.19479

[4] Meltem Sönmez Turan, Elaine Barker, John Kelsey, Kerry McKay, Mary Baish, and Mike Boyle, "Recommendation for the entropy sources used for random bit generation," National Institute of Standards and Technology, NIST Special Publication 800-90B, Jan. 2018. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-90B

[5] Bundesamt für Sicherheit in der Informationstechnik (BSI), "AIS 20/31: A proposal for: Functionality classes for random number generators, version 3.0," Bundesamt für Sicherheit in der Informationstechnik, Tech. Rep., Sep. 2024. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Certification/Interpretations/AIS_31_Functionality_classes_for_random_number_generators_e_2024.html

[6] Mitsuru Matsui, "Linear cryptanalysis method for DES cipher," in *Advances in Cryptology — EUROCRYPT '93*, ser. Lecture Notes in Computer Science, Tor Helleseth, Ed. Berlin, Heidelberg: Springer, 1994, vol. 765, pp. 386–397. [Online]. Available: https://doi.org/10.1007/3-540-48285-7_33

[7] Patrick Lacharme, "Post-processing functions for a biased physical random number generator," in *Fast Software Encryption, 15th International Workshop, FSE 2008, Lausanne, Switzerland, February 10–13, 2008, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 5086. Berlin, Heidelberg: Springer, 2008, pp. 334–342. [Online]. Available: https://doi.org/10.1007/978-3-540-71039-4_21

[8] Hongchao Zhou and Jehoshua Bruck, "Linear extractors for extracting randomness from noisy sources," in *Proceedings of the 2011 IEEE International Symposium on Information Theory, ISIT 2011, St. Petersburg, Russia, July 31–August 5, 2011*. IEEE, 2011, pp. 1738–1742. [Online]. Available: https://doi.org/10.1109/ISIT.2011.6033845

[9] Miloš Grujić and Ingrid Verbauwhede, "Optimizing linear correctors: A tight output min-entropy bound and selection technique," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 586–600, 2024. [Online]. Available: https://doi.org/10.1109/TIFS.2023.3326986

[10] Maciej Skórski, Francisco-Javier Soto, and Onur Günlü, "Exact bias of linear TRNG correctors – spectral approach," 2025. [Online]. Available: https://arxiv.org/abs/2509.26393

[11] Markku-Juhani O. Saarinen, "On entropy and bit patterns of ring oscillator jitter," in *2021 Asian Hardware Oriented Security and Trust Symposium (AsianHOST)*. IEEE, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/AsianHOST53231.2021.9699508

[12] Michele Bonnin, Fabio L. Traversa, and Fabrizio Bonani, "Colored noise in oscillators. phase-amplitude analysis and a method to avoid the Itô–Stratonovich dilemma," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 10, pp. 3917–3927, 2019. [Online]. Available: https://doi.org/10.1109/TCSI.2019.2914398

[13] Hongchao Zhou and Jehoshua Bruck, "Linear transformations for randomness extraction," 2012, extended version of the ISIT 2011 paper. [Online]. Available: https://arxiv.org/abs/1209.0732

[14] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone, *Handbook of Applied Cryptography*. Boca Raton, FL, USA: CRC Press, 1996.

[15] Robert B. Davies, "Exclusive OR (XOR) and hardware random number generators," Technical note, available at http://www.robertnz.net/pdf/xor2.pdf, 2002.

[16] Lawrence R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989. [Online]. Available: https://doi.org/10.1109/5.18626

[17] Andrei Simion, "Piling-up lemma for Markov chains," *Proceedings of the Romanian Academy, Series A*, vol. 10, no. 2, pp. 121–124, 2009.

[18] John von Neumann, "Various techniques used in connection with random digits," in *Monte Carlo Method*, ser. National Bureau of Standards Applied Mathematics Series, A. S. Householder, G. E. Forsythe, and H. H. Germond, Eds. Washington, D.C.: U.S. Government Printing Office, 1951, vol. 12, pp. 36–38.

[19] Peter Elias, "The efficient construction of an unbiased random sequence," *The Annals of Mathematical Statistics*, vol. 43, no. 3, pp. 865–870, Jun. 1972. [Online]. Available: https://doi.org/10.1214/aoms/1177692552

[20] Yuval Peres, "Iterating von Neumann's procedure for extracting random bits," *The Annals of Statistics*, vol. 20, no. 1, pp. 590–597, Mar. 1992. [Online]. Available: https://doi.org/10.1214/aos/1176348543

[21] Manuel Blum, "Independent unbiased coin flips from a correlated biased source: a finite state Markov chain," in *25th Annual Symposium on Foundations of Computer Science, West Palm Beach, Florida, USA, 24–26 October 1984*. IEEE Computer Society, 1984, pp. 425–433. [Online]. Available: https://doi.org/10.1109/SFCS.1984.715944

[22] Manuel Blum, "Independent unbiased coin flips from a correlated biased source—a finite state Markov chain," *Combinatorica*, vol. 6, no. 2, pp. 97–108, 1986. [Online]. Available: https://doi.org/10.1007/BF02579167

[23] Hongchao Zhou and Jehoshua Bruck, "Generalizing the Blum–Elias method for generating random bits from Markov chains," in *IEEE International Symposium on Information Theory, ISIT 2010, June 13–18, 2010, Austin, Texas, USA, Proceedings*. IEEE, 2010, pp. 1248–1252. [Online]. Available: https://doi.org/10.1109/ISIT.2010.5513679

[24] Hongchao Zhou and Jehoshua Bruck, "Efficient generation of random bits from finite state Markov chains," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2490–2506, Apr. 2012. [Online]. Available: https://doi.org/10.1109/TIT.2011.2175698

[25] Miklos Santha and Umesh V. Vazirani, "Generating quasi-random sequences from semi-random sources," *Journal of Computer and System Sciences*, vol. 33, no. 1, pp. 75–87, Aug. 1986. [Online]. Available: https://doi.org/10.1016/0022-0000(86)90044-9

[26] Salil P. Vadhan, "Pseudorandomness," *Foundations and Trends in Theoretical Computer Science*, vol. 7, no. 1–3, pp. 1–336, 2012. [Online]. Available: https://doi.org/10.1561/0400000010

[27] A. W. F. Edwards, "The meaning of binomial distribution," *Nature*, vol. 186, no. 4730, p. 1074, Jun. 1960. [Online]. Available: https://doi.org/10.1038/1861074a0

[28] Maciej Skórski and Johannes Mittmann, "Unpredictability from parity: Understanding linear correctors for markovian sources," Cryptology ePrint Archive, Paper 2026/2417, 2026. [Online]. Available: https://eprint.iacr.org/2026/2417

[29] Maciej Skorski, "markovian-parity-correctors: supplementary repository," GitHub repository, 2026. [Online]. Available: https://github.com/maciejskorski/markovian-parity-correctors

[30] A. F. Horadam, "Basic properties of a certain generalized sequence of numbers," *The Fibonacci Quarterly*, vol. 3, no. 3, pp. 161–176, Oct. 1965.

[31] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, UK: Cambridge University Press, 2013.

[32] Willi Meier and Othmar Staffelbach, "Fast correlation attacks on certain stream ciphers," *Journal of Cryptology*, vol. 1, no. 3, pp. 159–176, 1989. [Online]. Available: https://doi.org/10.1007/BF02252874

[33] Tosio Kato, *Perturbation Theory for Linear Operators*, 2nd ed., ser. Classics in Mathematics. Berlin, Heidelberg: Springer, 1995, vol. 132. [Online]. Available: https://doi.org/10.1007/978-3-642-66282-9

AI Disclosure

The authors used an AI assistant (Claude, Anthropic) for minor editorial polishing, related-work search, and numerical simulation code. The authors reviewed all AI-assisted output and take full responsibility for the article content.

APPENDIX

A. Proof of Theorem III.1

Proof. Fix n and let $w_j(\theta) = \mathbb{E}\big[t^{X_{j+1}+\cdots+X_n} \mid \theta_j = \theta\big]$, so $w_n \equiv \mathbf{1}$. Conditioning on θ_j with the Markov property gives $w_{j-1}(\theta) = \int \mathbf{P}^{(j)}(\theta, d\theta')\, t^{f(\theta')}\, w_j(\theta')$, i.e. $w_{j-1} = \mathbf{K}_t^{(j)} w_j$. Hence $w_0 = \mathbf{K}_t^{(1)} \cdots \mathbf{K}_t^{(n)} \mathbf{1}$ and $\mathbb{E}\, t^{X_1+\cdots+X_n} = \pi w_0$. For a fixed kernel this simplifies to $w_0 = \mathbf{K}_t^n \mathbf{1}$. □

B. Proof of Corollary III.2

Proof. Fix positions $i_1 < \cdots < i_q$ and let Q_j be the cumulative transition kernel from i_{j-1} to i_j with $i_0 := 0$ (the product of the $\mathbf{P}^{(\cdot)}$ over the gap), and $M_j = Q_j \mathbf{Z}$. Note $\|M_j\|_\infty \le 1$. Theorem III.1 applied to the chain $\theta_0, \theta_{i_1}, \ldots, \theta_{i_q}$ with kernels $Q_1, \ldots, Q_q$ gives $\mathsf{bias}(\bigoplus_{j=1}^q X_{i_j}) = \pi M_1 M_2 \cdots M_q \mathbf{1}$. For consecutive factors, the Markov property gives the conditional law of the middle state $\theta' = \theta_{i_j}$ given its neighbors $\theta = \theta_{i_{j-1}}$ and $\theta'' = \theta_{i_{j+1}}$, and therefore

$$(M_j M_{j+1})(\theta, d\theta'') = \sigma(\theta'')\, (Q_j Q_{j+1})(\theta, d\theta'')\, \mathbb{E}\big[\sigma(\theta') \mid \theta, \theta''\big].$$

By (1) this is at most $\beta_2\, (Q_j Q_{j+1})(\theta, d\theta'')$ in absolute value; integrating over θ'' gives $\|M_j M_{j+1}\|_\infty \le \beta_2$ (for consecutive positions this is the first claim). Group $M_2 \cdots M_q$ into $\lfloor (q-1)/2 \rfloor$ such pairs, plus a possible leftover factor of norm ≤ 1, so $\|M_2 \cdots M_q \mathbf{1}\|_\infty \le \beta_2^{\lfloor (q-1)/2 \rfloor}$. Since $\|M_1\|_\infty \le 1$ and π is a probability measure, the bound follows. □

C. Proof of Corollary III.4

Proof. From Corollary III.2: with $\mathsf{bias}(Y) = \mathbb{E}[(-1)^Y] = 1 - 2\mathbb{P}[Y=1]$, $\big|\mathbb{P}[Y=1] - \frac{1}{2}\big| = \frac{1}{2}|\,\mathsf{bias}(Y)| \le \frac{1}{2}\beta_2^{\lfloor (q-1)/2 \rfloor}$. □

D. Proof of Proposition III.5

Proof. For $n \ge 3$ let $A, C_2, \ldots, C_{n-1}$ be independent fair bits, with hidden states $\theta_1 = A$, $\theta_j = (A, C_j)$ for $1 < j < n$, and $\theta_n = A$, and bits $x_1 = x_n = A$, $x_j = C_j$ for $1 < j < n$ (for $n = 3$ this is (A, C, A)). The states form a first-order Markov chain (carry A, redraw C_j at each step, finally forget C), and each $x_j = f(\theta_j)$ is a function of θ_j, which depends on the past only through θ_{j-1}, so this is a hidden Markov source in the sense of [13], and a noisy hidden Markov source in the sense of [8] (Markov $y = \theta$, bits $x_j = I(y_j)$ with I the projection onto the emitted bit). For every $i < j < k$ with $1 < j < n$, the bit $x_j = C_j$ is independent of (θ_i, θ_k), which only involve A and C_i, C_k with $i, k \ne j$; hence $\mathbb{P}[x_j = 1 \mid \theta_i, \theta_k] = \frac{1}{2}$ and the hypothesis holds with $e = 0$, whether the triples are required to satisfy $i_3 < n$ (as in [13]; for $n = 3$ there are none) or only $i_3 \le n$ (as in [8]). But $x_1 \oplus x_n = 0$ identically, so with $q = 2$ the deviation $\big|\mathbb{P}[x_1 \oplus x_n = 1] - \frac{1}{2}\big| = \frac{1}{2}$ exceeds the claimed $\frac{1}{2}e^{1/2}$ for every $e < 1$. Their proof of Lemma 7 conditions on $\theta_{i_1}, \theta_{i_3}, \ldots$, which makes $x_{i_2}, x_{i_4}, \ldots$ independent with bias at most e each, but it does not separate the parity of q: for even q the last selected bit x_{i_q} has no right conditioning state, so the hypothesis does not apply to it and only $q/2 - 1$ factors are available (in the example, x_1, x_n are never the middle of a triple, so nothing controls them). □

E. Generating Function of Non-Homogeneous Order-m Chains

Corollary A.1 (Generating function, non-homogeneous order-m chains)**.** *Let $m \ge 1$ and let $X_{-m+1}, X_{-m+2}, \ldots$ be a (possibly time-inhomogeneous) order-m Markov chain on $\{0,1\}$. Let $W_i := (X_{i-m+1}, \ldots, X_i) \in \{0,1\}^m$ for $i \ge 0$, with initial distribution π and transition matrices $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \ldots$, and let $N_n := \sum_{i=1}^n X_i$ ($N_0 := 0$). Let $\mathbf{Z}_t := \mathrm{diag}\,\big(t^{\mathsf{last}(c)}\big)_c$ for a formal variable t, where $\mathsf{last}(c)$ is the last bit of c. Then*

$$\mathbb{E}(t^{N_n} \mid W_0 = c_0) = \Big(\mathbf{P}^{(1)} \mathbf{Z}_t \cdots \mathbf{P}^{(n)} \mathbf{Z}_t\, \mathbf{1}\Big)_{c_0},$$
$$\mathbb{E}(t^{N_n}) = \pi\, \mathbf{P}^{(1)} \mathbf{Z}_t \cdots \mathbf{P}^{(n)} \mathbf{Z}_t\, \mathbf{1}\,.$$

Proof of Corollary A.1. Recall $N_n = \sum_{i=1}^n X_i$ ($N_0 := 0$). Then $(N_n, W_n)_{n \ge 0}$ is a Markov chain on $\{0, \ldots, n\} \times \{0,1\}^m$. Define the probabilities

$$a_{c_0,c}^{(n)}(j) := \mathbb{P}(N_n = j, W_n = c \mid W_0 = c_0)\,.$$

Let t be a formal variable and define the matrix $\boldsymbol{A}^{(n)}(t) = (A_{c_0,c}^{(n)}(t))_{c_0,c}$ of generating functions by

$$A_{c_0,c}^{(n)}(t) := \sum_{j=0}^{n} a_{c_0,c}^{(n)}(j) \cdot t^j\,.$$

Then $\mathbb{E}(t^{N_n}) = \pi\, \boldsymbol{A}^{(n)}(t) \mathbf{1}$ is the probability generating function of N_n and $\mathbb{E}(t^{N_n} \mid W_0 = c_0) = (\boldsymbol{A}^{(n)}(t)\mathbf{1})_{c_0}$. We show by induction on n that

$$\boldsymbol{A}^{(n)}(t) = \mathbf{P}^{(1)} \mathbf{Z}_t \cdots \mathbf{P}^{(n)} \mathbf{Z}_t\,,$$

where $\mathbf{Z}_t$ is the diagonal matrix with $(\mathbf{Z}_t)_{c,c} = t^{\mathsf{last}(c)}$, where $\mathsf{last}(c) \in \{0,1\}$ denotes the last bit of c.

For $n = 0$, we have $A_{c_0,c}^{(0)}(t) = a_{c_0,c}^{(0)}(0) = [\,c_0 = c\,]$, hence $\boldsymbol{A}^{(0)}(t) = \mathbf{I}$ is the identity matrix (corresponding to the empty matrix product).

Now let $n \ge 1$. Since $X_n = \mathsf{last}(W_n)$, conditioning on W_{n-1} gives the recurrences

$$a_{c_0,c}^{(n)}(j) = \sum_{c'} a_{c_0,c'}^{(n-1)}(j) \cdot P_{c',c}^{(n)} \qquad \text{if } \mathsf{last}(c) = 0\,,$$
$$a_{c_0,c}^{(n)}(j) = \sum_{c'} a_{c_0,c'}^{(n-1)}(j-1) \cdot P_{c',c}^{(n)} \qquad \text{if } \mathsf{last}(c) = 1\,.$$

Therefore,

$$A_{c_0,c}^{(n)}(t) = \sum_{c'} A_{c_0,c'}^{(n-1)}(t) \cdot P_{c',c}^{(n)} \cdot t^{\mathsf{last}(c)}\,.$$

In matrix form, this reads $\boldsymbol{A}^{(n)}(t) = \boldsymbol{A}^{(n-1)}(t) \mathbf{P}^{(n)} \mathbf{Z}_t$, which proves the claim by induction.

In the homogeneous case $\mathbf{P}^{(i)} \equiv \mathbf{P}$, substituting $t = -1$ gives $\mathbf{Z}_{-1} = \mathbf{Z}$ and $\boldsymbol{A}^{(n)}(-1) = \mathbf{K}^n$, so $\mathsf{bias}(S_n) = \pi\, \mathbf{K}^n \mathbf{1}$, recovering Corollary III.6. □

F. Proof of Corollary III.6

Proof. We expand the expectation over all paths and recognize each path weight as a product of entries of $\mathbf{K}$. Let c_k be the context after k fresh bits, with c_0 the initial context, i.e. c_{k-1} shifted with x_k appended. The path $x_{1:n}$ has probability $\pi(c_0)\prod_{k=1}^{n}\mathbb{P}(x_k \mid c_{k-1})$, since x_k depends on the past only through c_{k-1}. The parity of the fresh bits is $\prod_{k=1}^{n}(-1)^{x_k}$, hence

$$\mathsf{bias}(S_n) = \sum_{c_0}\pi(c_0)\sum_{x_1,\dots,x_n}\prod_{k=1}^{n}(-1)^{x_k}\mathbb{P}\big(x_k \mid c_{k-1}\big).$$

Each factor is one entry of $\mathbf{K}$: $(-1)^{x_k}\mathbb{P}(x_k \mid c_{k-1}) = \mathbf{K}(c_{k-1}, c_k)$ (equivalently, $\mathbb{P}(x \mid s) = \frac{1+(-1)^x b_s}{2}$ gives $(-1)^x\mathbb{P}(x \mid s) = \frac{(-1)^x + b_s}{2}$). The map $(x_1,\dots,x_n) \mapsto (c_1,\dots,c_n)$ is a bijection onto the paths from c_0 with valid transitions, and $\mathbf{K}$ vanishes on all other pairs of contexts, so the inner sum equals $\sum_{c_1,\dots,c_n}\prod_{k=1}^{n}\mathbf{K}(c_{k-1},c_k) = (\mathbf{K}^n\mathbf{1})(c_0)$. Averaging over c_0 gives $\mathsf{bias}(S_n) = \pi\mathbf{K}^n\mathbf{1}$. The parity S_n^+ carries the additional sign $(-1)^{\mathrm{wt}(c_0)}$, which gives $\pi\mathbf{D}\mathbf{K}^n\mathbf{1}$. For time-dependent transition laws the same argument yields the product $\mathbf{K}_1\mathbf{K}_2\cdots\mathbf{K}_n$. The generating function is the homogeneous case of Corollary A.1, which is Theorem III.1 for the first-order chain of contexts. □

G. Proof of Corollary III.7

Warm-up ($m = 1$: one-bit state and identity emission). Let $S_n^+ := X_0 \oplus \cdots \oplus X_n$ (for $m = 1$ this is $S_n \oplus \mathrm{wt}(c_0)$) and $a_{i,j}^{(n)} := \mathbb{P}(S_n^+ = i, X_n = j)$. We prove by induction that

$$\begin{pmatrix} a_{0,0}^{(n)} - a_{1,0}^{(n)} & a_{0,1}^{(n)} - a_{1,1}^{(n)} \end{pmatrix} = \pi\mathbf{Z}(\mathbf{P}\mathbf{Z})^n.$$

The base case $n = 0$ holds by direct inspection; for $n > 0$:

$$\begin{aligned} a_{0,0}^{(n)} - a_{1,0}^{(n)} &= (a_{0,0}^{(n-1)} - a_{1,0}^{(n-1)})p_{0,0} + (a_{0,1}^{(n-1)} - a_{1,1}^{(n-1)})p_{1,0}, \\ a_{0,1}^{(n)} - a_{1,1}^{(n)} &= -(a_{0,0}^{(n-1)} - a_{1,0}^{(n-1)})p_{0,1} - (a_{0,1}^{(n-1)} - a_{1,1}^{(n-1)})p_{1,1}. \end{aligned}$$

This gives the recurrence with matrix $\mathbf{P}\mathbf{Z}$. Since $\mathsf{bias}(S_n^+) = (a_{0,0}^{(n)} - a_{1,0}^{(n)}) + (a_{0,1}^{(n)} - a_{1,1}^{(n)})$, this gives $\mathsf{bias}(X_0 \oplus S_n) = \pi\mathbf{Z}(\mathbf{P}\mathbf{Z})^n\mathbf{1}$. Taking $\pi = \delta_x$ and multiplying by $(-1)^x = \mathsf{bias}(X_0 \mid X_0 = x)$ yields $\mathsf{bias}(S_n \mid X_0 = x) = (\mathbf{K}^n\mathbf{1})_x$. □

Proof 1 (Induction, all orders). Here $X_i := f(\theta_i)$ for all $i \geq 0$, including X_0. Let $N_n^+ := X_0 + \cdots + X_n$ (integer sum, so $S_n^+ = N_n^+ \bmod 2$) and $a_{s,\theta}^{(n)} := \mathbb{P}(N_n^+ = s,\ \theta_n = \theta)$. Put $A_\theta^{(n)}(t) := \sum_s a_{s,\theta}^{(n)} t^s$; for $t = -1$ this is $a_{0,\theta}^{(n)} - a_{1,\theta}^{(n)}$ after reducing s mod 2. We prove by induction that

$$\big(A_\theta^{(n)}(t)\big)_\theta = \pi\mathbf{Z}_t\,(\mathbf{P}^{(1)}\mathbf{Z}_t)\cdots(\mathbf{P}^{(n)}\mathbf{Z}_t).$$

The base case $n = 0$ holds by inspection: $S_0^+ = f(\theta_0)$, so $A_\theta^{(0)}(t) = \pi_\theta t^{f(\theta)}$. For $n > 0$, conditioning on θ_{n-1} and using $X_n = f(\theta_n)$ gives:

$$a_{s,\theta'}^{(n)} = \sum_\theta a_{s-f(\theta'),\theta}^{(n-1)}\,\mathbf{P}^{(n)}(\theta,\theta'),$$

hence

$$A_{\theta'}^{(n)}(t) = t^{f(\theta')}\sum_\theta A_\theta^{(n-1)}(t)\,\mathbf{P}^{(n)}(\theta,\theta'),$$

which gives the recurrence with matrix $\mathbf{K}_t^{(n)} = \mathbf{P}^{(n)}\mathbf{Z}_t$. Since $\mathbb{E}\,t^{N_n^+} = \sum_\theta A_\theta^{(n)}(t)$, we get $\mathbb{E}\,t^{N_n^+} = \pi\mathbf{Z}_t(\mathbf{P}^{(1)}\mathbf{Z}_t)\cdots(\mathbf{P}^{(n)}\mathbf{Z}_t)\mathbf{1}$. Taking $\pi = \delta_\theta$ and dividing by $t^{f(\theta)}$ (an identity of polynomials in t) yields $\mathbb{E}\big[t^{N_n} \mid \theta_0 = \theta\big] = \big((\mathbf{P}^{(1)}\mathbf{Z}_t)\cdots(\mathbf{P}^{(n)}\mathbf{Z}_t)\mathbf{1}\big)_\theta$. At $t = -1$, $\mathbf{Z}_{-1} = \mathbf{Z} = \mathrm{diag}(\sigma)$ and $\mathbf{K}^{(i)} = \mathbf{P}^{(i)}\mathbf{Z}$, so $\mathsf{bias}(X_0 \oplus S_n) = \pi\mathbf{Z}\mathbf{K}^{(1)}\cdots\mathbf{K}^{(n)}\mathbf{1}$ and $\mathsf{bias}(S_n \mid \theta_0 = \theta) = (\mathbf{K}^{(1)}\cdots\mathbf{K}^{(n)}\mathbf{1})_\theta$. □

Proof 2 (Invariant Subspaces). Let $S_n^+ = \bigoplus_{k=0}^{n} X_k$ and $U_n = (X_n, S_n^+)$.

Step 1 (Markovian Augmentation). Augmenting the chain with the cumulative XOR preserves the Markov property: since $\mathbb{P}((x_n,y_n) \mid (x_{n-1},y_{n-1})) = \mathbb{P}(x_n \mid x_{n-1}) \cdot [\,y_n = y_{n-1} \oplus x_n\,]$, the process $\{U_n\}$ is Markovian with transition matrix

$$\mathbf{P}_U = \begin{pmatrix} \delta_1 & 0 & 0 & 1-\delta_1 \\ 0 & \delta_1 & 1-\delta_1 & 0 \\ 1-\delta_2 & 0 & 0 & \delta_2 \\ 0 & 1-\delta_2 & \delta_2 & 0 \end{pmatrix}, \tag{6}$$

where rows and columns are indexed by $(X_n, S_n^+) \in \{(0,0),(0,1),(1,0),(1,1)\}$ in lexicographic order.

Step 2 (Marginalization). For initial distribution π_U on U_0,

$$\mathsf{bias}(S_n^+) = \pi_U\mathbf{P}_U^n v, \quad v = [1,-1,1,-1]^\top. \tag{7}$$

Step 3 (Invariant Subspace Reduction). The 4×4 matrix powering above reduces to a 2×2 one. The subspace $\mathcal{H} = \{x \in \mathbb{R}^4 : x_1 + x_2 = 0,\ x_3 + x_4 = 0\}$ is invariant under $\mathbf{P}_U$. With $v_1 = [1,-1,0,0]^\top$, $v_2 = [0,0,1,-1]^\top$ spanning $\mathcal{H}$, and $T = [v_1, v_2, v_3, v_4]$ for $v_3 = [1,1,0,0]^\top$, $v_4 = [0,0,1,1]^\top$, we have the block decomposition

$$T^{-1}\mathbf{P}_U T = \begin{pmatrix} \delta_1 & \delta_1 - 1 & 0 & 0 \\ 1-\delta_2 & -\delta_2 & 0 & 0 \\ 0 & 0 & \delta_1 & 1-\delta_1 \\ 0 & 0 & 1-\delta_2 & \delta_2 \end{pmatrix}. \tag{8}$$

Since $v \in \mathcal{H}$, we can restrict to the upper-left 2×2 block, which is $\mathbf{K} = \mathbf{P}\mathbf{Z}$ with $\mathbf{Z} = \mathrm{diag}([1,-1])$, giving

$$\mathsf{bias}(S_n^+) = \pi_U[v_1\ v_2](\mathbf{P}\mathbf{Z})^n\begin{pmatrix}1\\1\end{pmatrix}, \qquad \pi_U = \mathbb{P}_{X_0,S_0^+}. \tag{9}$$

For $S_0^+ = X_0$ (the case of interest), $\pi_U = (\pi_0, 0, 0, \pi_1)$ and $\pi_U[v_1\ v_2] = (\pi_0, -\pi_1)$, which gives $\mathsf{bias}(S_n^+) = \pi\mathbf{Z}(\mathbf{P}\mathbf{Z})^n\mathbf{1}$; the conditional form $\mathsf{bias}(S_n \mid X_0 = x) = (\mathbf{K}^n\mathbf{1})_x$ follows as in the warm-up.

□

H. Proof of Theorem III.8

Proof. By Cayley–Hamilton, $\mathbf{K}^2 = \tau\mathbf{K} + \lambda\mathbf{I}$, so $\mathbf{K}^n = \tau\mathbf{K}^{n-1} + \lambda\mathbf{K}^{n-2}$ for $n \geq 2$. Multiplying by π on the left and $\mathbf{1}$ on the right and using $\mathsf{bias}(S_n) = \pi\mathbf{K}^n\mathbf{1}$ (Corollary III.6) gives the recursion. Put $a_n = \mathsf{bias}(S_n)$, and let F_{-1} be a formal symbol with $\lambda F_{-1} := 1$, so that $F_1 = \tau F_0 + \lambda F_{-1}$ holds without dividing by λ. Then $a_0 = 1 = \lambda F_{-1} + F_0 a_1$ and $a_1 = \lambda F_0 + F_1 a_1$, and both a_n and $\lambda F_{n-1} + F_n a_1$ satisfy the same second-order recursion, so they agree for all $n \geq 0$. □

I. Proof of Corollary III.9

Proof. Write $a_n = \mathsf{bias}(S_n)$, so $a_0 = 1$. If $\tau = 0$, Theorem III.8 gives $a_n = \lambda a_{n-2}$ for $n \geq 2$. Induction yields $a_{2k} = \lambda^k$ and $a_{2k+1} = \lambda^k a_1$, proving the displayed formulas and the even-step decay. For a stationary start, put $u = \mathbb{E}[(-1)^{X_0}] = \mathbb{E}[(-1)^{X_1}]$. The regression $\mathbb{E}[(-1)^{X_1} \mid X_0] = \lambda(-1)^{X_0}$ gives $u = \lambda u$, hence $u = a_1 = 0$ when $\lambda < 1$. At $\lambda = 1$ the chain is constant, the formulas still hold, but odd-step biases need not vanish.

If $\lambda = 0$, both rows of $\mathbf{P}$ coincide and $\mathbf{K}\mathbf{1} = \tau\mathbf{1}$, so $a_1 = \tau$ for every initial law of X_0. The recursion gives then $a_n = \tau a_{n-1} = \tau^n$ for $n \geq 1$. Equivalently, the emitted bits are independent with bias τ, recovering the Piling-up Lemma. □

J. Proof of Theorem III.10

Proof. By Cayley–Hamilton's formula, $\mathbf{K}^n = F_n\mathbf{K} + \lambda F_{n-1}\mathbf{I}$, and $(\mathbf{K}\mathbf{1})_x = \tau + (-1)^x\lambda$. Hence

$$\mathsf{bias}(S_n \mid X_0 = x) = F_{n+1} + (-1)^x \lambda F_n,$$

and therefore

$$\max_x |\,\mathsf{bias}(S_n \mid X_0 = x)| = |F_{n+1}| + |\lambda| \cdot |F_n|. \tag{10}$$

If $\rho = 0$, then $\tau = \lambda = 0$, so $F_{n+1} = 0$ for $n \geq 1$ and both sides vanish; let $\rho > 0$ below. Let η_1, η_2 be the eigenvalues ordered such that $\rho = |\eta_1| \geq |\eta_2|$. Let $h = \eta_2/\eta_1$ and denote $\Sigma_n = \sum_{k=0}^n h^k$. Then $F_{n+1} = \sum_{k=0}^n \eta_1^k \eta_2^{n-k} = \eta_1^n \Sigma_n$, so $|F_{n+1}| = \rho^n |\Sigma_n|$ and $|F_n| = \rho^{n-1}|\Sigma_{n-1}|$. Also $\lambda = -\eta_1\eta_2 = -\eta_1^2 h$ so $|\lambda| = \rho^2|h|$ and $\Sigma_n = 1 + h\Sigma_{n-1}$; hence $|\lambda||F_n| = \rho^{n+1}|\Sigma_n - 1|$ and

$$\max_x |\,\mathsf{bias}(S_n \mid X_0 = x)| = \rho^n|\Sigma_n| + \rho^{n+1}|\Sigma_n - 1|.$$

The two-sided bound follows from $|\Sigma_n - 1| \leq |\Sigma_n| + 1$ and $|\Sigma_n| + |\Sigma_n - 1| \geq 1$:

$$\frac{\rho^n(\rho + |\Sigma_n|)}{1+\rho} \leq \max_x |\,\mathsf{bias}(S_n \mid X_0 = x)| \leq (1+\rho)\rho^n(\rho + |\Sigma_n|).$$

Regimes (discussion after the theorem). Let $M_n = \max_x |\,\mathsf{bias}(S_n \mid X_0 = x)|/\rho^n = |\Sigma_n| + \rho|\Sigma_n - 1|$; for a fixed start the ratio is at most M_n. If $h = 1$, then $\Sigma_n = n+1$ and $M_n = (1+\rho)n + 1$. If $h = e^{i\theta} \neq 1$, then $\Sigma_n = \frac{1-h^{n+1}}{1-h}$, so $|\Sigma_n| = \frac{|\sin((n+1)\theta/2)|}{|\sin(\theta/2)|} \leq \frac{1}{|\sin(\theta/2)|}$, and $|\Sigma_n - 1| = |h\Sigma_{n-1}| = |\Sigma_{n-1}|$ obeys the same bound; hence $M_n \leq (1+\rho)/|\sin(\theta/2)|$. If $|h| < 1$, then $\Sigma_n \to \frac{1}{1-h}$ and $M_n \to \frac{1+\rho|h|}{|1-h|}$. □

K. Proof of Theorem III.11

Proof. Since $\mathsf{bias}(S_n) = \pi\mathbf{K}^n\mathbf{1}$ is linear in the row vector π (the law of X_0), its maximum over the probability simplex is attained at a vertex, i.e. a point mass; this gives the first equality in (5), $\max_\pi |\,\mathsf{bias}(S_n)| = \|\mathbf{K}^n\mathbf{1}\|_\infty$.

Each row of $\mathbf{K}$ has absolute entries summing to $\mathbb{P}(0 \mid x) + \mathbb{P}(1 \mid x) = 1$, so $\|\mathbf{K}\|_\infty = 1$, and in particular $\|\mathbf{K}\mathbf{1}\|_\infty = \max\{|b_0|, |b_1|\} = \beta$. By Cayley–Hamilton, $\mathbf{K}^2 = \tau\mathbf{K} + \lambda\mathbf{I}$ with $\tau = \operatorname{tr}\mathbf{K}$, so

$$\|\mathbf{K}^2\|_\infty \leq |\tau|\,\|\mathbf{K}\|_\infty + |\lambda|\,\|\mathbf{I}\|_\infty = |\tau| + |\lambda| = \beta.$$

Then submultiplicativity of $\|\cdot\|_\infty$, for $n = 2m$:

$$\|\mathbf{K}^n\mathbf{1}\|_\infty \leq \|\mathbf{K}^2\|_\infty^m \|\mathbf{1}\|_\infty \leq \beta^m,$$

and for $n = 2m+1$:

$$\|\mathbf{K}^n\mathbf{1}\|_\infty = \|\mathbf{K}^{2m}(\mathbf{K}\mathbf{1})\|_\infty \leq \|\mathbf{K}^2\|_\infty^m \|\mathbf{K}\mathbf{1}\|_\infty \leq \beta^{m+1}.$$

Both cases give $\|\mathbf{K}^n\mathbf{1}\|_\infty \leq \beta^{\lceil n/2 \rceil}$.

For equality, take $\delta_1 = \delta_2 = \frac{1\pm\beta}{2}$ and X_0 deterministic: then $\tau = 0$ and $\mathbf{K}^2 = \lambda\mathbf{I}$ with $|\lambda| = \beta$, so every entry of $\mathbf{K}^n\mathbf{1}$ has modulus exactly $\beta^{\lceil n/2 \rceil}$. (For non-homogeneous chains every two-step product $\mathbf{K}^{(i)}\mathbf{K}^{(i+1)}$ still contracts by β: the entries e_0, e_1 of its row x satisfy $e_0 - e_1 = (\mathbf{K}^{(i)}\mathbf{1})_x$ and $e_0 + e_1 = (\mathbf{K}^{(i)}\mathbf{K}^{(i+1)}\mathbf{1})_x$, both at most β in absolute value, so $\|\mathbf{K}^{(i)}\mathbf{K}^{(i+1)}\|_\infty \leq \beta$ and the pairing argument gives $\beta^{\lceil n/2 \rceil}$ for $\mathbf{K}^{(1)} \cdots \mathbf{K}^{(n)}\mathbf{1}$.) □

L. Proof of Corollary III.13

Proof. Let $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$ and put $t = (b-1)n$. By the Markov property and homogeneity,

$$\mathbb{E}[(-1)^{Y_b} \mid \mathcal{F}_t] = (\mathbf{K}^n\mathbf{1})_{X_t}.$$

By Theorem III.11, the absolute value is at most $\beta^{\lceil n/2 \rceil}$ for every start state, including when $b = 1$ and $t = 0$. Since $Y_{<b}$ is measurable with respect to $\mathcal{F}_t$, the tower property and conditional triangle inequality give

$$\begin{aligned}\big|\mathbb{E}[(-1)^{Y_b} \mid Y_{<b}]\big| &\leq \mathbb{E}\big[\,\big|\mathbb{E}[(-1)^{Y_b} \mid \mathcal{F}_t]\big|\,\big|Y_{<b}\big] \\ &\leq \beta^{\lceil n/2 \rceil}.\end{aligned}$$

For non-homogeneous chains, Corollary A.2 replaces $(\mathbf{K}^n\mathbf{1})_{X_t}$ by $(\mathbf{K}^{(t+1)} \cdots \mathbf{K}^{(t+n)}\mathbf{1})_{X_t}$, and the remark in the proof of Theorem III.11 gives the same bound. □

M. Proof of Proposition III.14

Proof. Each block of $m+1$ consecutive bits $X_{i-m}, \ldots, X_i$ ($i \geq 1$) has parity ε_i, which is independent of the past and has bias β. Tile $X_{-m+1}, \ldots, X_n$ by such blocks from the right end; this leaves $r = (n+m) \bmod (m+1) \leq m$ leading bits, which lie in the initial context and are 0. The $\lfloor (n+m)/(m+1) \rfloor = \lceil n/(m+1) \rceil$ blocks have independent parities ε_i, so the bias is $\beta^{\lceil n/(m+1) \rceil}$. □

N. Proof of Proposition III.15

Proof. We will split $\mathbf{K} = \mathbf{K}_0 + B$ into the nilpotent kernel $\mathbf{K}_0$ of a fair source, $\mathbf{K}_0^{m+1} = 0$, and a bias part B with $\|B\|_\infty \le \beta$. Use the induced norm $\|A\|_\infty = \max_s \sum_{s'} |A(s,s')|$. On a valid transition $s \to s' = (s_2, \dots, s_m, x)$,

$$\mathbf{K}(s,s') = (-1)^x \mathbb{P}(x \mid s) = \tfrac{(-1)^x + b_s}{2} = \mathbf{K}_0(s,s') + B(s,s'),$$

with $\mathbf{K}_0(s,s') = (-1)^x/2$ and $B(s,s') = b_s/2$ (both zero off valid transitions). Each state has two successors, so $\|\mathbf{K}_0\|_\infty = 1$ and $\|B\|_\infty \le \beta$. The absolute entries of a row of $\mathbf{K}$ are the two probabilities $\mathbb{P}(x \mid s)$, so $\|\mathbf{K}\|_\infty = 1$.

Let $L = m+1$. A path of L steps emitting $x_1, \dots, x_L$ has weight $2^{-L}(-1)^{x_1 + \cdots + x_L}$ and ends in $s'' = (x_2, \dots, x_L)$: after m steps the memory holds only emitted bits, and at step L the first one, x_1, drops out. For fixed s'' only x_1 is free, so

$$\mathbf{K}_0^L(s,s'') = 2^{-L}(-1)^{\mathrm{wt}(s'')} \sum_{x_1 \in \{0,1\}} (-1)^{x_1} = 0.$$

Hence $\mathbf{K}_0^L = 0$. Expanding $(\mathbf{K}_0 + B)^L$ into ordered products, the pure $\mathbf{K}_0$ term vanishes, and a product with $j \ge 1$ factors B has norm at most β^j ($\binom{L}{j}$ such products), so

$$\|\mathbf{K}^L\|_\infty \le \sum_{j=1}^{L} \binom{L}{j} \beta^j = (1+\beta)^L - 1 = \gamma.$$

For $n = dL + r$, $0 \le r < L$: $\|\mathbf{K}^n \mathbf{1}\|_\infty \le \|\mathbf{K}^L\|_\infty^d \|\mathbf{K}^r\|_\infty \|\mathbf{1}\|_\infty \le \gamma^{\lfloor n/L \rfloor}$, as $\|\mathbf{K}^r\|_\infty \le 1$. Since $\mathsf{bias}(S_n) = \pi \mathbf{K}^n \mathbf{1}$ and $\mathsf{bias}(S_n^+) = \pi \mathbf{D} \mathbf{K}^n \mathbf{1}$ with π a probability vector and $\|\mathbf{D}\|_\infty = 1$, the claim follows. □

O. Proof of Proposition III.16

Proof. Use the splitting $\mathbf{K} = \mathbf{K}_0 + B$ from the proof of Proposition III.15: $\mathbf{K}_0^L = 0$, $\|\mathbf{K}_0\|_\infty = 1$ and $\|B\|_\infty \le \beta$. The case $\beta = 0$ gives $B = 0$, so $\mathbf{K} = \mathbf{K}_0$, $\mathbf{K}^L = 0$ and $\rho(\mathbf{K}) = 0$; assume $\beta > 0$.

Step 1 (Resolvent of the nilpotent part). Let $\mathbf{K} f = \eta f$ with $f \ne 0$; for $\eta = 0$ there is nothing to prove. Then $(\eta \mathbf{I} - \mathbf{K}_0) f = B f$. Since $\mathbf{K}_0^L = 0$, the Neumann series [31], [33] terminates:

$$(\eta \mathbf{I} - \mathbf{K}_0)^{-1} = \sum_{k=0}^{L-1} \eta^{-k-1} \mathbf{K}_0^k,$$

as multiplying by $\eta \mathbf{I} - \mathbf{K}_0$ telescopes to $\mathbf{I} - \eta^{-L} \mathbf{K}_0^L = \mathbf{I}$. Hence

$$f = \sum_{k=0}^{L-1} \eta^{-k-1} \mathbf{K}_0^k B f.$$

Step 2 (Norm bound). By submultiplicativity $\|\mathbf{K}_0^k\|_\infty \le 1$, so

$$\|f\|_\infty \le \beta \sum_{k=1}^{L} |\eta|^{-k} \|f\|_\infty.$$

Dividing by $\|f\|_\infty$ and multiplying by $|\eta|^L$ gives $|\eta|^L \le \beta(1 + |\eta| + \cdots + |\eta|^{L-1})$. In other words, η is not an eigenvalue when $\beta \sum_{k=1}^{L} |\eta|^{-k} < 1$, since then $\eta \mathbf{I} - \mathbf{K} = (\eta \mathbf{I} - \mathbf{K}_0)\big(\mathbf{I} - (\eta \mathbf{I} - \mathbf{K}_0)^{-1} B\big)$ is invertible.

Step 3 (The scalar root). The map $r \mapsto 1 - \beta \sum_{k=1}^{L} r^{-k}$ is strictly increasing on $(0, \infty)$, from $-\infty$ to 1. It therefore has a unique zero r_*, and Step 2 says it is ≤ 0 at $r = |\eta|$. Hence $|\eta| \le r_*$ and $\rho(\mathbf{K}) \le r_*$.

Step 4 (Explicit bounds). Also $\rho(\mathbf{K}) \le \|\mathbf{K}\|_\infty = 1$ [31]. If $r_* \le 1$, then $r_*^L = \beta \sum_{j<L} r_*^j \le L\beta$. If $r_* > 1$, the monotone map of Step 3 is negative at $r = 1$, i.e. $L\beta > 1$. In both cases $\rho(\mathbf{K}) \le \min\{1, (L\beta)^{1/L}\}$. The function $t \mapsto t^{1/t}$ increases on $(0, e)$ and decreases on (e, ∞), and $2^{1/2} < 3^{1/3}$, so $\max_{L \in \mathbb{N}} L^{1/L} = 3^{1/3} < 1.443$. Finally, if $r_* < 1$ then $r_*^L = \beta \sum_{j<L} r_*^j \le \beta/(1 - r_*)$. Since $r_* \le (L\beta)^{1/L}$, this gives $\rho(\mathbf{K})^L \le \beta(1 + O(\beta^{1/L}))$. □

P. Proof of Theorem III.18

Proof. Upper bound: let $g = \mathbb{E}[(-1)^{X_j} \mid X_{<j}]$, so $|g| \le \beta$. Then $\big|\mathsf{bias}\big(\bigoplus_{i \in A} X_i\big)\big| = \big|\mathbb{E}\big[(-1)^{\bigoplus_{i \in A \setminus \{j\}} X_i} g\big]\big| \le \mathbb{E}|g| \le \beta$. Lower bound: let $\mathbb{P}(X_i = 0 \mid X_{<i}) = \frac{1}{2}$ for $i \ne j$ and $\mathbb{P}(X_j = 0 \mid X_{<j}) = \frac{1}{2}\big(1 + \beta(-1)^{\bigoplus_{i \in A \setminus \{j\}} X_i}\big) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$. Then $\mathbb{E}[(-1)^{X_j} \mid X_{<j}] = \beta(-1)^{\bigoplus_{i \in A \setminus \{j\}} X_i}$ and the parity has bias β. □

Q. Proof of Corollary III.19

Proof. Apply Theorem III.18 to the support A of a nonzero row: the corresponding output bit Y_1 has $|\mathsf{bias}(Y_1)| = \beta$, so $\Delta_{\mathsf{SD}}(Y) \ge \Delta_{\mathsf{SD}}(Y_1) = \beta/2$, and some value y_1 has $\mathbb{P}(Y_1 = y_1) = \frac{1+\beta}{2}$; summing over the 2^{k-1} completions, some y has $\mathbb{P}(Y = y) \ge 2^{-(k-1)} \frac{1+\beta}{2}$. □

R. Order-m Chains: Numerical Evidence

Numerical evidence: maximizing $\rho(\mathbf{K})$ over all sign patterns $b_s = \pm\beta$ and random $b_s \in [-\beta, \beta]$ gives $\sup \rho(\mathbf{K})^{m+1}/\beta = 1.000$ for $m = 2, 3$ at $\beta = 0.02$, attained by $b_s = \pm\beta(-1)^{\mathrm{wt}(s)}$, which is the noisy LFSR. The threshold β_m is necessary: for $m = 2$ and $\beta = 0.5$ the chain $b = \beta(1, -1, 1, 1)$ has $\rho(\mathbf{K}) = 0.831 > \beta^{1/3} = 0.794$, where ρ is the largest root of $\eta^3 = \beta\eta^2 + \frac{\beta(1-\beta)}{2}(\eta + 1)$ (the crossover is at $\beta \approx 0.2$). Moreover the bound $|\mathsf{bias}(S_n)| \le \beta^{\lceil n/(m+1) \rceil}$, exact for $m = 1$, fails for $m \ge 2$: for $m = 2$, $n = 4$ the maximum is about $1.5\beta^2$ for small β.

S. Non-Homogeneous Chains, $m = 1$

Corollary A.2 (Non-homogeneous extension, $m = 1$)**.** *For $m = 1$ ($W_i = X_i$) and $X_0 \sim \pi$, Corollary A.1 at $t = -1$ gives, with the product ordered left to right,*

$$\begin{aligned} \mathsf{bias}(S_n) &= \pi \prod_{i=1}^{n} (\mathbf{P}^{(i)} \mathbf{Z})\, \mathbf{1}, \\ \mathsf{bias}(X_0 \oplus S_n) &= \pi \mathbf{Z} \prod_{i=1}^{n} (\mathbf{P}^{(i)} \mathbf{Z})\, \mathbf{1}. \end{aligned} \tag{11}$$

Proof of Corollary A.2. Set $m = 1$ and $t = -1$ in Corollary A.1. Then $W_i = X_i$ and $\mathbf{Z}_{-1} = \mathrm{diag}(1, -1) = \mathbf{Z}$, giving the first identity. Conditionally on $X_0 = x$, including X_0 in the parity multiplies its sign by $(-1)^x$. Averaging over $X_0 \sim \pi$ therefore replaces the initial row vector π by $\pi \mathbf{Z}$, giving the second identity with the same ordered product. □