

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 reduce the bias of random bits via parity (XOR), but their guarantees are known mostly for independent bits, an assumption violated by many physical entropy sources. We give a *fully general, exact formula* for the parity bias of dependent bits from a (hidden, possibly non-homogeneous) Markov chain, as a product of twisted transfer kernels. Under next-bit unpredictability, i.e. conditional bias at most β , the decay is governed by β itself: *sharp* for a one-bit context, $|\text{bias}| \leq \beta^{\lceil n/2 \rceil}$, and *near-optimal* for an m -bit context, within a factor $3^{1/3} \approx 1.44$ of the best possible per-bit rate, which is at least $\beta^{1/(m+1)}$ (attained by a noisy LFSR) and asymptotically equal to it as $\beta \rightarrow 0$. Against the best previous bound (Zhou and Bruck, ISIT 2011) we *correct the exponent and tighten the bound on n bits by a factor of about $2^{\lfloor (n-1)/2 \rfloor} / \beta$* , with a matching construction. The decay is non-monotone, so a corrector should be certified at every operating length, or via the envelope $\beta^{\lceil n/2 \rceil}$. Finally, *bounded memory is necessary*: if unpredictability is imposed on the full history, XOR-ing does not reduce the bias at all.

Index Terms—randomness extraction, bias correction, Markov chains, XOR correctors, 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].

Yet rigorous guarantees, needed for security, are almost entirely limited to independent bits: the classical Piling-up Lemma [4] gives the geometric bias-decay rate for independent coins, and the follow-up literature largely keeps that assumption [5]–[8]. 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 [9], and oscillator noise is often exponentially correlated (Ornstein–Uhlenbeck) [10], 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 ($\theta = X$, $f = \text{id}$), 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 $S_n = X_1 \oplus \dots \oplus X_n$ and particularly their bias $\text{bias}(S_n) = \mathbb{E}(-1)^{S_n}$. The source quality is quantified by its *next-bit bias* β , $|\mathbb{E}[(-1)^{X_i} \mid \theta_{\leq i-1}]| \leq \beta$, 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, Theorem A.1):

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

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 $\text{bias}(S_n) = \pi \mathbf{K}^n \mathbf{1}$ with $\mathbf{K} = \mathbf{P} \mathbf{Z}_{-1}$ (Theorem III.6). For hidden sources, a two-sided next-bit bias β_2 on the middle bit given both neighbouring states gives $|\text{bias}(\bigoplus_j X_{i_j})| \leq \beta_2^{\lfloor (q-1)/2 \rfloor}$ for any q selected positions (Theorem III.2); this corrects the exponent in Zhou and Bruck's [11] Lemma 7 (Theorems III.4 and III.5).

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 $\tau = \text{tr}(\mathbf{K})$, $\lambda = -\det(\mathbf{K})$) gives a Fibonacci-type recursion with an explicit Binet formula, also for coinciding eigenvalues (Theorems III.8 and III.10):

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

The bias vanishes at the geometric rate $\rho(\mathbf{K})^n$, up to a factor of $n + 1$. If the next-bit bias is at most β , then $|\text{bias}(S_n)| \leq \beta^{\lceil n/2 \rceil}$, 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 (Theorem III.15)

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

The parity bias of a noisy linear-feedback-shift-register (LFSR [12]) decays at exactly $\beta^{1/(m+1)}$ per bit (Theorem III.13), so no bound in terms of β can beat this rate; ours is within a factor $(m+1)^{1/(m+1)} \leq 3^{1/3} \approx 1.44$ of it for every m , hence of the best possible rate. Moreover, $\rho(\mathbf{K})^{m+1} \leq \beta(1 + O(\beta^{1/(m+1)}))$ (Theorem III.15), 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 [13]; linear correctors were formalized by Dichtl [1] and analyzed for independent bits by Lacharme [5] and Zhou and Bruck [6], and, most recently, by Skórski, Soto and Günlü [8], who give exact formulas and near-tight distance-to-uniformity bounds via code weight enumerators. All these results reduce to the parity biases $\text{bias}(c \cdot X)$ of the input, which factor for independent bits by the Piling-up Lemma [4]; the missing ingredient for dependent inputs is precisely the parity bias, which we compute for Markov sources.

Dependent inputs. Zhou and Bruck [6], [11] give the best previous bound for Markov sources: for hidden Markov sources, including order-one chains with $\Pr(x_i | x_{i-1}) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$, their Lemma 9 (Lemma 7 in the extended version) claims $|\text{bias}| \leq e^{(q-1)/2}$ for q bits, with $e = 2\beta/(1+\beta^2)$. We improve this in three ways: (i) the exponent is false in general (Theorem III.5), and we prove the correct one, $\lfloor (q-1)/2 \rfloor$ (Theorems III.2 and III.4); (ii) for binary chains the base $2\beta/(1+\beta^2)$ improves to β and the exponent $(q-1)/2$ to $\lceil n/2 \rceil$ on n consecutive bits, a gain of about $2^{(n-1)/2}/\beta$ (Theorem III.3); (iii) the resulting bound $\beta^{\lceil n/2 \rceil}$ for n consecutive bits is sharp (Theorem III.11).

Simion [14] 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.¹ On the negative side, Kukorelly [15] shows that the piling-up approximation can be far off for arbitrary dependence. Extraction from Markov sources by non-linear,

¹The closed forms in cases 2 and 4 of [14, Sec. 2] are inaccurate: for $x_1 = x_2$ they carry a spurious factor 2^{-m} , and for coinciding eigenvalues the coefficient $2n - 3$ should be n ; the correct expressions follow from Theorem III.8.

TABLE I
WORST-CASE BIAS $|\text{bias}(S_n)|$ OF $S_n = X_1 \oplus \dots \oplus X_n$ WHEN THE NEXT-BIT BIAS IS AT MOST β ; FOR ORDER m , THE DECAY RATE PER BIT.

Memory	Reference	Result
none (i.i.d.)	Piling-up Lemma [4]	β^n
order 1	Theorem III.11	$\beta^{\lceil n/2 \rceil}$ (sharp)
order m	Theorems III.13 and III.15	$\text{rate} \in [\beta^{\frac{1}{m+1}}, 3^{1/3} \beta^{\frac{1}{m+1}}]$
unbounded	Theorem III.17	β for every n

variable-length procedures (von Neumann, Elias, Peres, Blum) is classical [16]–[22]; we analyze fixed-rate linear correctors, preferred in hardware.

Markovian counting processes. Edwards [23] 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 $\text{bias}(Z) = \Pr(Z = 0) - \Pr(Z = 1) = \mathbb{E}[(-1)^Z]$; it determines the law of Z . For independent X_j , parity multiplies biases (Piling-up Lemma [4]): $\text{bias}(\bigoplus_{j \in J} X_j) = \prod_{j \in J} \text{bias}(X_j)$.

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

$$\beta := \sup_{n \geq 1, \theta_{\leq n-1}} |\mathbb{E}[(-1)^{X_n} | \theta_{\leq n-1}]|,$$

that is, $\Pr(X_n = x | \theta_{\leq n-1}) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$; $\beta = 0$ 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.

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 bias of the XOR of n bits with next-bit bias at most β . Proofs and secondary results are in the appendix.

Intuition. XOR-ing independent bits multiplies their biases. With memory, a bit may encode the parity of its predecessors: one twisted step does not contract ($\|\mathbf{K}\|_\infty = 1$), and for $m = 1$ only a pair of steps does, by a factor β . An order- m chain can turn the parity of $m + 1$ consecutive bits into a single fresh noise bit, hence one factor β per block of $m + 1$ bits, and with unbounded memory the last bit can absorb the parity of all earlier ones, leaving bias β for every n .

A. Generating-Function Identity

Let $(\theta_i)_{i \geq 0}$ 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 $t \in \mathbb{C}$, the *twisted kernel* is $\mathbf{K}_t^{(i)} = \mathbf{P}^{(i)} \mathbf{Z}_t$ with $\mathbf{Z}_t = \text{diag}(t^f)$: one step of the chain, weighted by t^{X_i} of the state reached. At $t = -1$ we write $\mathbf{K}^{(i)} = \mathbf{K}_{-1}^{(i)}$, $\mathbf{Z} = \mathbf{Z}_{-1}$ and $\sigma = (-1)^f$, so that $\mathbf{K}^{(i)}(\theta, d\theta') = \mathbf{P}^{(i)}(\theta, d\theta') \sigma(\theta')$.

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

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

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

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

$$|\mathbb{E}[(-1)^{X_j} \mid \theta_i, \theta_k]| \leq \beta_2. \quad (1)$$

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

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

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 neighbours, so they contract by β_2 , and every second selected position costs one factor β_2 .

Remark III.3 (Relation to Zhou and Bruck [11]). Their model is the hidden Markov source $X_i = f(\theta_i)$, with the noise hypothesis (1) (parameter e) imposed for all $i < j < k \leq n$ in Lemma 9 of the ISIT paper [6] (and $k < n$ in Lemma 7 of the extended version [11]); it is the hypothesis of Theorem III.2. Under this hypothesis, (1) charges every pair of neighbours at its worst case, giving $\beta_2^{\lfloor (q-1)/2 \rfloor}$ (it conditions on both neighbours, so the last selected bit of an even-size set is not covered), whereas they claim the stronger $\beta_2^{(q-1)/2}$, which Theorem III.5 shows is false in general. For a first-order binary chain ($\theta = X$, $\Pr(X_i \mid X_{i-1}) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$) 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 : for $\beta = 0.1$ and $n = 9$ it is $1.5 \cdot 10^{-3}$ against 10^{-5} . Our bound is attained for every n (Theorem III.11), theirs for none, 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 [11] that $|\Pr[x_{i_2} = 1 \mid \theta_{i_1}, \theta_{i_3}] - \frac{1}{2}| \leq \frac{e}{2}$ for all $i_1 < i_2 < i_3$, Theorem III.2 with $\beta_2 = e$ gives, for any selected positions $i_1 < \dots < i_q$,*

$$\left| \Pr \left[\bigoplus_{\ell=1}^q x_{i_\ell} = 1 \right] - \frac{1}{2} \right| \leq \frac{1}{2} e^{\lfloor (q-1)/2 \rfloor}.$$

Proposition III.5 (Lemma 9 of [6] (Lemma 7 of [11]) 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 [11] with $e = 0$, yet $x_1 \oplus x_3 \equiv 0$, so $|\Pr[x_1 \oplus x_3 = 1] - \frac{1}{2}| = \frac{1}{2} > \frac{1}{2} e^{1/2}$ for every $e < 1$, contradicting their lemma for $q = 2$. The same holds for every length $n \geq 3$ and under either index convention (Section 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 $s = (s_1, \dots, s_m)$ with $f(s) = s_m$ (the last symbol), so Theorem III.1 applies. Its kernel is $\mathbf{P}(s, s') = \Pr(x \mid s)$ if $s' = (s_2, \dots, s_m, x)$ and 0 otherwise. With $\mathbf{Z}_t = \text{diag}(t^{s_m})$ 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 $S_n = \sum_{i=1}^n X_i$ (at $t = -1$ only its parity matters),*

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

At $t = -1$, for $S_n = X_1 \oplus \dots \oplus X_n$,

$$\text{bias}(S_n) = \pi \mathbf{K}^n \mathbf{1},$$

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

and for the parity $S_n^+ = S_n \oplus \text{wt}(c_0)$ that includes the initial context, with $\mathbf{D} = \text{diag}((-1)^{\text{wt}(s)})$,

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

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 specializes 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,*

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

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

$$\text{bias}(X_0 \oplus S_n) = \pi \mathbf{Z} \mathbf{K}^n \mathbf{1} \quad (X_0 \sim \pi).$$

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

We now use a more convenient parametrization of $\mathbf{P}$ by the balance τ and the correlation λ , in which the next-bit bias is

$|\tau| + |\lambda|$ and the parity bias obeys a Fibonacci-type recursion. With the next-bit biases $b_x = \mathbb{E}[(-1)^{X_{n+1}} | X_n = x]$, i.e. $b_0 = 2\delta_1 - 1$ and $b_1 = 1 - 2\delta_2$, put

$$\begin{aligned}\tau &= \text{tr } \mathbf{K} = \delta_1 - \delta_2 = \frac{b_0 + b_1}{2}, \\ \lambda &= -\det \mathbf{K} = \delta_1 + \delta_2 - 1 = \frac{b_0 - b_1}{2},\end{aligned}\quad (2)$$

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, from X_1). *Let $S_n = \bigoplus_{i=1}^n X_i$ (with $S_0 := 0$, so $\text{bias}(S_0) = 1$). Then*

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

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

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 for τ, λ .

The proof uses Cayley–Hamilton’s formula $\mathbf{K}^2 = \tau \mathbf{K} + \lambda \mathbf{I}$, 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

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

So $|\text{bias}(S_n)| = |\lambda|^{n/2}$ for even n ; for a stationary start, $\text{bias}(S_1) = 0$ and the odd-step bias vanishes identically.

Memoryless chain, $\lambda = 0$. The recursion collapses to $\text{bias}(S_n) = \tau \text{bias}(S_{n-1})$, giving $\text{bias}(S_n) = \tau^n \text{bias}(S_0)$ — 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, assuming $\rho > 0$ (otherwise $\mathbf{K}^2 = 0$ and all biases vanish for $n \geq 1$), set $h := \eta_2/\eta_1$. Let $\Sigma_n = \sum_{i=0}^n h^i$. Then for $S_n = \bigoplus_{i=1}^n X_i$ and every $n \geq 1$,*

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

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

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

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

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

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

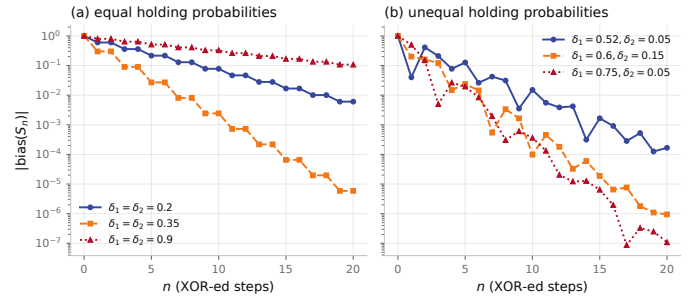


Fig. 1. Decay of $|\text{bias}(S_n)|$ with the number of XOR-ed steps, for chains with equal (left) and unequal (right) holding probabilities (δ_1, δ_2) , starting from $X_0 = 0$.

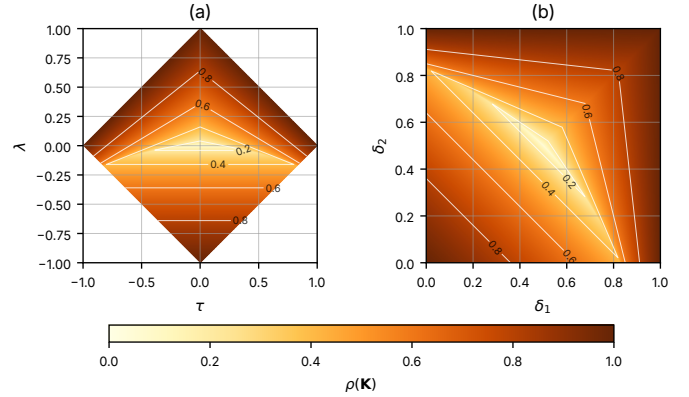


Fig. 2. Spectral radius $\rho(\mathbf{K})$ in terms of the two parametrizations: balance-correlation (τ, λ) (a) and holding probabilities (δ_1, δ_2) (b).

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, $\beta^{\lceil n/2 \rceil}$ (Figure 1 shows the oscillating decay); e.g. for $\beta = 0.1$ and $n = 8$ bits, 10^{-8} becomes 10^{-4} . The bound depends only on β , i.e. on the worst-case conditional min-entropy $1 - \log_2(1 + \beta)$.

The decay is not monotone: in panel (b), $|\text{bias}(S_n)|$ drops to 0.04 at $n = 1$ and rebounds to 0.41 at $n = 2$. Hence a guarantee at one length does not extend to others: a corrector that looks safe at one n can fail at the next, so one should certify the whole trajectory $n \mapsto \text{bias}(S_n)$ over all operating lengths, or rely on the envelope $\beta^{\lceil n/2 \rceil}$, as in cryptographic evaluation.

Figures 2 and 3 show how ρ and the leading constant depend on the parameters.

C. General Order- m Chains: Rate Bounds

For general memory m , Theorem III.6 no longer reduces to a simple closed-form rate; we bracket the per-bit rate, governed by the spectral radius of $\mathbf{K}$, between a lower and an upper bound.

Proposition III.13 (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*

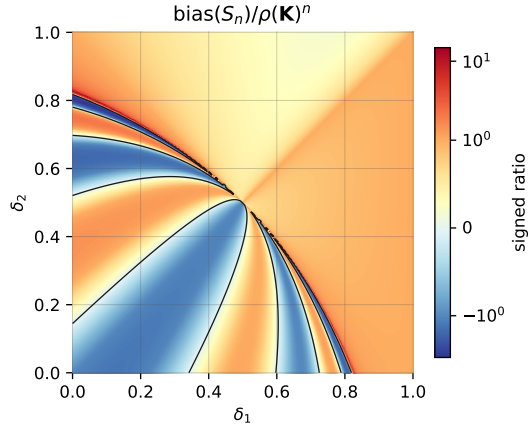


Fig. 3. Leading constant $C_n = \text{bias}(S_n)/\rho(\mathbf{K})^n$ (signed), for $n = 10$ and $X_0 = 0$, over (δ_1, δ_2) .

β (a noisy LFSR [12] with all-ones taps) and initial context $X_{-m+1} = \dots = X_0 = 0$. Then $\text{bias}(S_n) = \beta^{\lceil n/(m+1) \rceil}$.

Here the XOR of any $m+1$ consecutive outputs equals one fresh noise bit ε_i , so each block of $m+1$ bits costs one factor β ; unlike in correlation attacks [24], where noise corrupts the observed output, it enters here through the feedback.

Proposition III.14 (Elementary upper bound). *For every order- m chain and every $n \geq 0$, $|\text{bias}(S_n)| \leq \gamma^{\lceil n/(m+1) \rceil}$ with $\gamma = (1 + \beta)^{m+1} - 1 \approx (m+1)\beta$, and the same bound holds for S_n^+ .*

Idea: split $\mathbf{K} = \mathbf{K}_0 + B$ with $\mathbf{K}_0^{m+1} = 0$ and $\|B\|_\infty \leq \beta$, and expand $\mathbf{K}^{m+1}$ binomially.

Proposition III.15 (Spectral radius via nilpotence). *Let $L = m+1$. For every order- m chain, every eigenvalue η of $\mathbf{K}$ satisfies*

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

Hence $\rho(\mathbf{K}) \leq r_*$, the unique nonnegative root of $r^L = \beta(1 + r + \dots + r^{L-1})$ (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 $\beta \rightarrow 0$.

With Theorem III.13 (noisy LFSR, bias decaying at exactly $\beta^{1/L}$ per bit), the optimal per-bit rate lies between $\beta^{1/L}$ and $(L\beta)^{1/L}$: within a factor $L^{1/L} \leq 3^{1/3} \approx 1.44$ for every m and β ; as $\beta \rightarrow 0$ the refined bound $\rho(\mathbf{K})^L \leq \beta(1 + O(\beta^{1/L}))$ of Theorem III.15 makes the rate asymptotically exact. The proof (Section L) uses the resolvent of the nilpotent part of $\mathbf{K}$ in place of the binomial expansion behind Theorem III.14, and plays the role of Cayley–Hamilton for $m = 1$.

Conjecture III.16. *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 Theorem III.13.*

Numerics over all sign patterns $b_s = \pm\beta$ give $\sup \rho(\mathbf{K})^{m+1}/\beta = 1.000$ for $m = 2, 3$, attained by the noisy LFSR (Section P, which also explains why the threshold β_m is needed).

D. Finite Memory is Necessary: Santha–Vazirani Barrier

Call a process a *Santha–Vazirani source* 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; let $\text{SV}(\beta)$ be the class of such sources. Parity filters then fail.

Theorem III.17 (Parity filters fail on Santha–Vazirani sources). *Let $A \subset \mathbb{N}$ be finite and nonempty, and let $j = \max A$. Then*

$$\max_{(X_k) \in \text{SV}(\beta)} \left| \text{bias} \left(\bigoplus_{i \in A} X_i \right) \right| = \beta,$$

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 $\sup |\text{bias}(S_n)| = \beta$ for every $n \geq 1$.

Corollary III.18 (Linear correctors do not help). *Let $G \in \mathbb{F}_2^{k \times n}$ have a nonzero row and $Y = GX$, and let $\Delta_{\text{SD}}(Y) = \frac{1}{2} \sum_y |\Pr(Y = y) - 2^{-k}|$ be its statistical distance from uniform. For every $\beta \in (0, 1)$ there is a Santha–Vazirani source with bias β and $\Delta_{\text{SD}}(Y) \geq \beta/2$ and $\max_y \Pr(Y = y) \geq 2^{-k}(1 + \beta)$, i.e. $H_\infty(Y) \leq k - \log_2(1 + \beta)$. Hence the min-entropy deficiency of the output is at least that of a single raw bit, for every n and every G .*

The extremal source copies the parity of the earlier bits of A into X_j , so it needs memory spanning A ; for order- m chains gains come only from disjoint blocks of $m+1$ bits, one factor β per block (Theorems III.13 and III.15); 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**, $\text{bias}(S_n) = \pi \mathbf{K}^{(1)} \dots \mathbf{K}^{(n)} \mathbf{1}$, 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** (certify every operating length, or use the envelope), and a rate **within a factor $3^{1/3} \approx 1.44$** of the best possible one for order- m chains. We **correct the exponent** of the lemma of Zhou and Bruck and **tighten their bound on n bits by about $2^{\lfloor (n-1)/2 \rfloor} / \beta$** , with a matching construction; and **bounded memory is necessary**: under full-history unpredictability, XOR-ing gives no gain.

Next comes total variation distance for multi-bit outputs $Y = GX$, via the standard XOR lemma: our formula already gives the parity biases of arbitrary supports (via the subsampled chain), and what remains is to bound them sharply, aggregate them over a code, and settle the sharp rate for higher-order chains (Theorem III.16).

REFERENCES

- [1] M. Dichtl, “Bad and good ways of post-processing biased physical random numbers,” in *Fast Software Encryption (FSE 2007)*, ser. LNCS, vol. 4593. Springer, 2007, pp. 137–152.
- [2] C. Klein, O. Cret, and A. Suciu, “Design and implementation of a high quality and high throughput TRNG in FPGA,” 2009.
- [3] X. Guo, F. Lin, J. Lin, Z. Song, Y. Luo, Q. Wang, and Y. Guo, “Parallel and real-time post-processing for quantum random number generators,” 2024.
- [4] M. Matsui, “Linear Cryptanalysis Method for DES Cipher,” in *Advances in Cryptology — EUROCRYPT ’93*, G. Goos, J. Hartmanis, and T. Hellese, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, vol. 765, pp. 386–397. [Online]. Available: http://link.springer.com/10.1007/3-540-48285-7_33
- [5] P. Lacharme, “Post-processing functions for a biased physical random number generator,” in *International Workshop on Fast Software Encryption*. Springer, 2008, pp. 334–342.
- [6] H. Zhou and J. Bruck, “Linear extractors for extracting randomness from noisy sources,” in *2011 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2011, pp. 1738–1742.
- [7] M. Grujić and I. 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://ieeexplore.ieee.org/document/10292710/>
- [8] M. Skórski, F.-J. Soto, and O. Günlü, “Exact bias of linear TRNG correctors – spectral approach,” 2025.
- [9] M.-J. O. Saarinen, “On entropy and bit patterns of ring oscillator jitter,” in *Proc. 6th Asian Hardware Oriented Security and Trust Symposium (AsianHOST)*. IEEE, 2021, pp. 1–6.
- [10] M. Bonnin, F. L. Traversa, and F. Bonani, “Colored noise in oscillators. phase-amplitude analysis and a method to avoid the Itô–Stratonovich dilemma,” *IEEE Trans. Circuits Syst. I: Regul. Pap.*, vol. 66, no. 10, pp. 3917–3927, 2019.
- [11] H. Zhou and J. Bruck, “Linear transformations for randomness extraction,” 2012, extended version of the ISIT 2011 paper.
- [12] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. CRC Press, 1996.
- [13] R. B. Davies, “Exclusive OR (XOR) and hardware random number generators,” <http://www.robertnz.net/pdf/xor2.pdf>, 2002.
- [14] A. Simion, “Piling-up lemma for markov chains,” in *Proceedings of the Romanian Academy, Series A*, vol. 10, no. 2, 2009, pp. 121–124.
- [15] Z. Kukorely, “The piling-up lemma and dependent random variables,” in *Cryptography and Coding, 7th IMA International Conference, Cirencester, UK, December 20–22, 1999, Proceedings, 1999*, pp. 186–190. [Online]. Available: https://doi.org/10.1007/3-540-46665-7_22
- [16] J. von Neumann, “Various techniques used in connection with random digits,” in *Monte Carlo Method*, A. Householder, G. Forsythe, and H. Germond, Eds. Washington, D.C.: U.S. Government Printing Office: National Bureau of Standards Applied Mathematics Series, 12, 1951, pp. 36–38.
- [17] P. Elias, “The efficient construction of an unbiased random sequence,” *Ann. Math. Statist.*, vol. 43, no. 3, pp. 865–870, 06 1972. [Online]. Available: <https://doi.org/10.1214/aoms/1177692552>
- [18] Y. Peres, “Iterating von neumann’s procedure for extracting random bits,” *Ann. Statist.*, vol. 20, no. 1, pp. 590–597, 03 1992. [Online]. Available: <https://doi.org/10.1214/aos/1176348543>
- [19] M. 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, 1984*, pp. 425–433. [Online]. Available: <https://doi.org/10.1109/SFCS.1984.715944>
- [20] —, “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>
- [21] H. Zhou and J. 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, 2010*, pp. 1248–1252. [Online]. Available: <https://doi.org/10.1109/ISIT.2010.5513679>
- [22] —, “Efficient generation of random bits from finite state markov chains,” *IEEE Trans. Information Theory*, vol. 58, no. 4, pp. 2490–2506, 2012. [Online]. Available: <https://doi.org/10.1109/TIT.2011.2175698>
- [23] A. W. F. Edwards, “The meaning of binomial distribution,” *Nature*, vol. 186, no. 4730, p. 1074, 1960.
- [24] W. Meier and O. Staffelbach, “Fast correlation attacks on certain stream ciphers,” *Journal of Cryptology*, vol. 1, no. 3, pp. 159–176, 1989.
- [25] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2013.

USE OF AI TOOLS

The authors used an AI assistant (Claude, Anthropic) to polish the writing, to search the related work, and to write code for numerical simulations. The authors reviewed AI-assisted output and take full responsibility for the article content.

APPENDIX

A. Proof of Theorem III.1

Proof. Let $v_k(\theta) = \mathbb{E}[t^{X_1+\dots+X_k} \mid \theta_0 = \theta]$ for the chain started at $\theta_0 = \theta$ (with the kernels shifted accordingly). Then $v_0 = 1$, and conditioning on θ_1 with the Markov property gives $v_{k+1}(\theta) = \int \mathbf{P}^{(1)}(\theta, d\theta') t^{f(\theta')} v_k^{\text{shift}}(\theta')$, i.e. $v_{k+1} = \mathbf{K}_t^{(1)} v_k^{\text{shift}}$. Hence $v_n = \mathbf{K}_t^{(1)} \dots \mathbf{K}_t^{(n)} \mathbf{1}$ and $\mathbb{E} t^{X_1+\dots+X_n} = \pi v_n$. For a fixed kernel this reads $v_k = \mathbf{K}_t^k \mathbf{1}$. $\square$

B. Proof of Theorem III.2

Proof. Fix positions $i_1 < \dots < i_q$ and let Q_j be the transition kernel of the subsampled chain from i_{j-1} to i_j (the product of the $\mathbf{P}^{(\cdot)}$ over the gap), and $M_j = Q_j \mathbf{Z}$. Theorem III.1 applied to the subsampled chain gives $\text{bias}(\bigoplus_j X_{i_j}) = \mathbb{E}[\sigma(\theta_{i_1})(M_2 \dots M_q \mathbf{1})(\theta_{i_1})]$. For consecutive factors, the Markov property gives the conditional law of the middle state $\theta' = \theta_{i_j}$ given its neighbours $\theta = \theta_{i_{j-1}}$ and $\theta'' = \theta_{i_{j+1}}$, and therefore

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

By (1) this is at most $\beta_2(Q_j Q_{j+1})(\theta, d\theta'')$ in absolute value; integrating over θ'' gives $\|M_j M_{j+1}\|_\infty \leq \beta_2$. Group $M_2 \dots M_q$ into $\lfloor (q-1)/2 \rfloor$ such pairs, plus at most one leftover factor of norm ≤ 1 . Since $|\sigma| = 1$, the bound follows. $\square$

C. Proof of Theorem III.4

Proof. Immediate from Theorem III.2: with $\text{bias}(Y) = \mathbb{E}[(-1)^Y] = 1 - 2\Pr[Y = 1]$, $|\Pr[Y = 1] - \frac{1}{2}| = \frac{1}{2} |\text{bias}(Y)| \leq \frac{1}{2} \beta_2^{\lfloor (q-1)/2 \rfloor}$. $\square$

D. Proof of Theorem III.5

Proof. For $n \geq 3$ let $A, C_2, \dots, C_{n-1}$ 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 depends on the past only through θ_{j-1} , so this is a hidden Markov source in the sense of [11]. 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 $i, k \neq j$; hence $\Pr[x_j = 1 \mid \theta_i, \theta_k] = \frac{1}{2}$ and the hypothesis holds with $e = 0$, whether the triples are required to satisfy $i_3 < n$ (as in [11]; for $n = 3$ there are none) or only $i_3 \leq n$. But

$x_1 \oplus x_n = 0$ identically, so with $q = 2$ the deviation $|\Pr[x_1 \oplus x_n = 1] - \frac{1}{2}| = \frac{1}{2}$ exceeds the claimed $\frac{1}{2}e^{1/2}$ for every $e < 1$. The proof of their Lemma 7 conditions on $\theta_{i_1}, \theta_{i_3}, \dots$ and uses the hypothesis to make $x_{i_2}, x_{i_4}, \dots$ fair; for even q the last selected bit has no right neighbour among the selected positions, and the end bits x_1, x_n are never the middle of a triple, so nothing controls them. $\square$

E. Proof of Theorem 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 $\pi(c_0) \prod_{k=1}^n \Pr(x_k | c_{k-1})$, 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

$$\text{bias}(S_n) = \sum_{c_0} \pi(c_0) \sum_{x_1, \dots, x_n} \prod_{k=1}^n (-1)^{x_k} \Pr(x_k | c_{k-1}).$$

Each factor is one entry of $\mathbf{K}$: $(-1)^{x_k} \Pr(x_k | c_{k-1}) = \mathbf{K}(c_{k-1}, c_k)$ (equivalently, $\Pr(x | s) = \frac{1+(-1)^x b_s}{2}$ gives $(-1)^x \Pr(x | s) = \frac{(-1)^x + b_s}{2}$). 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 $\text{bias}(S_n) = \pi \mathbf{K}^n \mathbf{1}$. The parity S_n^+ carries the additional sign $(-1)^{\text{wt}(c_0)}$, which gives $\pi \mathbf{D} \mathbf{K}^n \mathbf{1}$. For time-dependent transition laws the same argument yields the product $\mathbf{K}_1 \mathbf{K}_2 \dots \mathbf{K}_n$. The generating function is the homogeneous case of Theorem A.1, which is Theorem III.1 for the first-order chain of contexts. $\square$

F. Proof of Theorem III.7

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

$$(a_{0,0}^{(n)} - a_{1,0}^{(n)} \quad a_{0,1}^{(n)} - a_{1,1}^{(n)}) = \pi \mathbf{Z}(\mathbf{P} \mathbf{Z})^n.$$

The base case $n = 0$ holds by 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 $\text{bias}(S_n^+) = (a_{0,0}^{(n)} - a_{1,0}^{(n)}) + (a_{0,1}^{(n)} - a_{1,1}^{(n)})$, this gives $\text{bias}(X_0 \oplus S_n) = \pi \mathbf{Z}(\mathbf{P} \mathbf{Z})^n \mathbf{1}$. Taking $\pi = \delta_x$ and multiplying by $(-1)^x = \text{bias}(X_0 | X_0 = x)$ yields $\text{bias}(S_n | X_0 = x) = (\mathbf{K}^n \mathbf{1})_x$. $\square$

Proof 1 (Induction, all orders). Let $S_n^+ := X_0 + \dots + X_n$ (integer sum; its parity is the S_n^+ above) and $a_{s,\theta}^{(n)} := \Pr(S_n^+ = s, \theta_n = \theta)$. 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

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

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 $\mathbb{E} t^{S_n^+} = \sum_\theta A_\theta^{(n)}(t)$, we get $\mathbb{E} t^{X_0 + S_n} = \pi \mathbf{Z}_t (\mathbf{P}^{(1)} \mathbf{Z}_t) \dots (\mathbf{P}^{(n)} \mathbf{Z}_t) \mathbf{1}$. Taking $\pi = \delta_\theta$ and dividing by $t^{f(\theta)}$ (an identity of polynomials in t) yields $\mathbb{E}[t^{S_n} | \theta_0 = \theta] = ((\mathbf{P}^{(1)} \mathbf{Z}_t) \dots (\mathbf{P}^{(n)} \mathbf{Z}_t) \mathbf{1})_\theta$. At $t = -1$, $\mathbf{Z}_{-1} = \mathbf{Z} = \text{diag}(\sigma)$ and $\mathbf{K}^{(i)} = \mathbf{P}^{(i)} \mathbf{Z}$, so $\text{bias}(X_0 \oplus S_n) = \pi \mathbf{Z} \mathbf{K}^{(1)} \dots \mathbf{K}^{(n)} \mathbf{1}$ and $\text{bias}(S_n | \theta_0 = \theta) = (\mathbf{K}^{(1)} \dots \mathbf{K}^{(n)} \mathbf{1})_\theta$. $\square$

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 $\Pr((x_n, y_n) | (x_{n-1}, y_{n-1})) = \Pr(x_n | x_{n-1}) \cdot [y_n = y_{n-1} \oplus x_n]$, the process $\{U_n\}$ is Markovian with transition matrix

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

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 ,

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

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

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

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

$$\text{bias}(S_n^+) = \pi_U [v_1 \ v_2] (\mathbf{P} \mathbf{Z})^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \pi_U = \Pr_{X_0, S_0^+}. \quad (9)$$

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 the claim. $\square$

G. 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 $\text{bias}(S_n) = \pi\mathbf{K}^n\mathbf{1}$ (Theorem III.6) gives the recursion. Put $a_n = \text{bias}(S_n)$, 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 $n \geq 0$. $\square$

H. Proof of Theorem III.10

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

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

and hence also

$$\max_x |\text{bias}(S_n | X_0 = x)| = |F_{n+1}| + |\lambda| \cdot |F_n|, \quad (10)$$

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

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

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

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

$\square$

I. Proof of Theorem III.11

Proof. Since $\text{bias}(S_n) = \pi\mathbf{K}^n\mathbf{1}$ 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), $\max_\pi |\text{bias}(S_n)| = \|\mathbf{K}^n\mathbf{1}\|_\infty$.

Each row of $\mathbf{K}$ has absolute entries summing to $\Pr(0 | x) + \Pr(1 | x) = 1$, 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 $\tau = \text{tr } \mathbf{K}$, 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 $\beta^{\lceil n/2 \rceil}$. Finally, for $X_0 \oplus S_n$ the bias is $\pi\mathbf{Z}\mathbf{K}^n\mathbf{1}$ with $\|\pi\mathbf{Z}\mathbf{1}\|_1 = 1$, so the same bounds apply. $\square$

J. Proof of Theorem III.13

Proof. Each block of $m + 1$ consecutive bits $X_{i-m}, \dots, X_i$ ($i \geq 1$) has parity ε_i , which is independent of the past and has bias β . Tile $X_{-m+1}, \dots, X_n$ 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 $\beta^{\lceil n/(m+1) \rceil}$. $\square$

K. Proof of Theorem III.14

Proof. Use the induced norm $\|A\|_\infty = \max_s \sum_{s'} |A(s, s')|$. On a valid transition $s \rightarrow s' = (s_2, \dots, s_m, x)$,

$$\mathbf{K}(s, s') = (-1)^x \Pr(x | s) = \frac{(-1)^x + b_s}{2} = \mathbf{K}_0(s, s') + B(s, s'),$$

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 $\|B\|_\infty \leq \beta$. The absolute entries of a row of $\mathbf{K}$ are the two probabilities $\Pr(x | s)$, so $\|\mathbf{K}\|_\infty = 1$.

Let $L = m + 1$. An entry of $\mathbf{K}_0^L$ sums over paths emitting L bits. The final context fixes only the last m bits, so for fixed start and end contexts the first emitted bit is free; the two choices have equal magnitude and opposite sign and cancel. 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 $j \geq 1$ factors B has norm at most $\beta^j \binom{L}{j}$ such products), so

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

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

L. Proof of Theorem III.15

Proof. Use the splitting $\mathbf{K} = \mathbf{K}_0 + B$ from the proof of Theorem III.14: $\mathbf{K}_0^L = 0$, $\|\mathbf{K}_0\|_\infty = 1$ and $\|B\|_\infty \leq \beta$. 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 $f \neq 0$; for $\eta = 0$ there is nothing to prove. Then $(\eta\mathbf{I} - \mathbf{K}_0)f = Bf$. Since $\mathbf{K}_0^L = 0$, the Neumann series [25] 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

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

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

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

Dividing by $\|f\|_\infty$ and multiplying by $|\eta|^L$ gives $|\eta|^L \leq \beta(1 + |\eta| + \dots + |\eta|^{L-1})$. Equivalently, $\eta\mathbf{I} - \mathbf{K} = (\eta\mathbf{I} - \mathbf{K}_0)(\mathbf{I} - (\eta\mathbf{I} - \mathbf{K}_0)^{-1}\mathbf{B})$ is invertible whenever $\beta \sum_{k=1}^L |\eta|^{-k} < 1$.

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 $|\eta| \leq r_*$ and $\rho(\mathbf{K}) \leq r_*$.

Step 4 (Explicit bounds). Also $\rho(\mathbf{K}) \leq \|\mathbf{K}\|_\infty = 1$. If $r_* \leq 1$, then $r_*^L = \beta \sum_{j < L} r_*^j \leq L\beta$. If $r_* > 1$, the monotone map of Step 3 is negative at $r = 1$, i.e. $L\beta > 1$. In both cases $\rho(\mathbf{K}) \leq \min\{1, (L\beta)^{1/L}\}$. 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 $r_*^L = \beta \sum_{j < L} r_*^j \leq \beta/(1 - r_*)$. Since $r_* \leq (L\beta)^{1/L}$, this gives $\rho(\mathbf{K})^L \leq \beta(1 + O(\beta^{1/L}))$. $\square$

M. Proof of Theorem III.17

Proof. Upper bound: let $g = \mathbb{E}[(-1)^{X_j} \mid X_{<j}]$, so $|g| \leq \beta$. Then $|\text{bias}(\bigoplus_{i \in A} X_i)| = |\mathbb{E}[(-1)^{\bigoplus_{i \in A \setminus \{j\}} X_i} g]| \leq \mathbb{E}|g| \leq \beta$. Lower bound: let $\Pr(X_i = 0 \mid X_{<i}) = \frac{1}{2}$ for $i \neq j$ and $\Pr(X_j = 0 \mid X_{<j}) = \frac{1}{2}(1 + \beta(-1)^{\bigoplus_{i \in A \setminus \{j\}} X_i}) \in [\frac{1-\beta}{2}, \frac{1+\beta}{2}]$. 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 . $\square$

N. Proof of Theorem III.18

Proof. Apply Theorem III.17 to the support A of a nonzero row: the corresponding output bit Y_1 has $|\text{bias}(Y_1)| = \beta$, so $\Delta_{\text{SD}}(Y) \geq \Delta_{\text{SD}}(Y_1) = \beta/2$, and some value y_1 has $\Pr(Y_1 = y_1) = \frac{1+\beta}{2}$; summing over the 2^{k-1} completions, some y has $\Pr(Y = y) \geq 2^{-(k-1)} \frac{1+\beta}{2}$. $\square$

O. Further Bounds

P. 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 $b_s = \pm\beta(-1)^{\text{wt}(s)}$, 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 $|\text{bias}(S_n)| \leq \beta^{\lceil n/(m+1) \rceil}$, exact for $m = 1$, fails for $m \geq 2$: for $m = 2$, $n = 4$ the maximum is about $1.5\beta^2$ for small β .

Q. Non-Homogeneous Chains

Corollary A.1 (Generating function, non-homogeneous order- m chains). *Let $m \geq 1$ and let $X_{-m+1}, X_{-m+2}, \dots$ be a (possibly time-inhomogeneous) order- m Markov chain on $\{0, 1\}$. Let $W_i := (X_{i-m+1}, \dots, X_i) \in \{0, 1\}^m$ for $i \geq 0$, with initial distribution π and transition matrices $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots$, and let $S_n := \sum_{i=1}^n X_i$ ($S_0 := 0$). Let $\mathbf{Z}_t := \text{diag}(t^{\text{last}(c)})_c$ for a formal variable t , where $\text{last}(c)$ is the last bit of c . Then*

$$\mathbb{E}(t^{S_n} \mid W_0 = c_0) = \left(\mathbf{P}^{(1)} \mathbf{Z}_t \cdots \mathbf{P}^{(n)} \mathbf{Z}_t \mathbf{1} \right)_{c_0},$$

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

Proof of Theorem A.1. More generally, allow $X_{-m+1}, X_{-m+2}, \dots$ to be a (possibly time-inhomogeneous) order- m Markov chain on $\{0, 1\}$. Denote $W_i := (X_{i-m+1}, \dots, X_i)$ taking values in $\{0, 1\}^m$ for $i \geq 0$. Let $\pi = (\pi_{c_0})_{c_0}$ be the distribution of W_0 and denote the transition probability of W_i given W_{i-1} by $\mathbf{P}^{(i)} = (P_{c_{i-1}, c_i}^{(i)})_{c_{i-1}, c_i}$. That is, $W_0, W_1, \dots$ is a first-order Markov chain on $\{0, 1\}^m$ with initial distribution π and transition probabilities $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots$.

Recall $S_n = \sum_{i=1}^n X_i$ ($S_0 := 0$). Then $(S_n, W_n)_{n \geq 0}$ is a Markov chain on $\{0, \dots, n\} \times \{0, 1\}^m$. Define the probabilities

$$a_{c_0, c}^{(n)}(j) := \Pr(S_n = j, W_n = c \mid W_0 = c_0).$$

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

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

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

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

where $\mathbf{Z}_t$ is the diagonal matrix with $(\mathbf{Z}_t)_{c, c} = t^{\text{last}(c)}$, where $\text{last}(c) \in \{0, 1\}$ 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 $\mathbf{A}^{(0)}(t) = \mathbf{I}$ is the identity matrix (corresponding to the empty matrix product).

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

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

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

Therefore,

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

In matrix form, this reads $\mathbf{A}^{(n)}(t) = \mathbf{A}^{(n-1)}(t) \mathbf{P}^{(n)} \mathbf{Z}_t$, 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 $\mathbf{A}^{(n)}(-1) = \mathbf{K}^n$, so $\text{bias}(S_n) = \pi \mathbf{K}^n \mathbf{1}$, recovering Theorem III.6. $\square$

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

$$\text{bias}(S_n) = \pi \prod_{i=1}^n (\mathbf{P}^{(i)} \mathbf{Z}) \mathbf{1}, \tag{11}$$

$$\text{bias}(X_0 \oplus S_n) = \pi \mathbf{Z} \prod_{i=1}^n (\mathbf{P}^{(i)} \mathbf{Z}) \mathbf{1}.$$