Finite Persistent Memory in Exact Sampling: Countable Rigidity and Algebraic Memory Bounds

Scott Schubr
Luminary Oddities

October 22, 2026
Archive edition 2

Abstract

An exact sampler reads independent fair bits and emits an iid discrete law, retaining a fixed finite state between outputs. Temporary work and per-round prefix-code depth are unrestricted; an external iteration counter counts as retained information. We prove that zero asymptotic expected path loss is possible on finite or countable support exactly when every positive mass is a reciprocal integer power of two. Nonnegative finite-path loss also covers infinite entropy. The necessity proof reduces countably many output kernels to finitely many integer types and an exact positive rational normalization. For finite targets, a signed-charge cycle inequality handles cancelling arithmetic resonances. Combined with classical logarithmic-form estimates, it gives a polynomial loss lower bound in the number of states for every fixed algebraic nonexceptional law. The known upper construction then yields persistent memory of order $\log(1/\varepsilon)$ bits. Both bounds transfer uniformly to computable conditional dyadic refinements, including infinite-entropy examples. An explicit resonant law illustrates the signed-charge mechanism.

1 Introduction and relation to earlier work

Resources and result hierarchy The resource is what remains between outputs: M persistent states, or $\lceil \log_2 M \rceil$ retained bits. Temporary workspace and the round depth are unrestricted; an external iteration counter is not free persistent information. Three conclusions have distinct scopes: zero-loss classification covers countable laws; logarithmic memory order concerns each fixed finite algebraic nonexceptional law; and the countable refinement upper construction uses the supplied computable conditional codes. Individual reciprocal powers of two are a narrower condition than arbitrary dyadic-rational masses.

Random-bit efficiency and retained memory are distinct resources in exact sampling. A sampler may perform arbitrarily long computation within a round while retaining only one of finitely many states after an output. Draper and Saad [1] formulate this persistent-state model, construct a state-information potential, and obtain cycle-based space–entropy lower bounds. Their Section 9 asks whether any target outside the reciprocal-power class can achieve zero asymptotic loss with finitely many states. Their quantitative Theorem 18 applies under a nonnegative logarithmic-representation hypothesis that excludes certain cancelling arithmetic relations.

We resolve the zero-loss question for every countable discrete law in the model below. For finite algebraic laws we also remove the quantitative resonance restriction: a signed arithmetic charge, followed by an iid block-fluctuation estimate, forces positive loss even when some zero-gap product contains every symbol. The quantitative bounds extend to countable conditional dyadic

refinements, uniformly in their tail complexity. The known upper constructions of Draper–Saad [2] and Draper–Harris–Saad [3] are explicitly retained as prior work.

The state-information and cycle framework is inherited from [1]; the new arguments address its countable endpoint and cancelling arithmetic relations. Classical prefix coding supplies the reciprocal-power construction. Section 8 compares the relevant coding and arithmetic antecedents under their precise resource conventions.

All logarithms are base two unless written ln. A *reciprocal-power law* means that every positive mass is $2^{-\ell}$ for some integer $\ell \geq 0$; this is narrower than merely requiring dyadic-rational masses.

2 Model, path loss and a state potential

Fix a law $p = (p_x)_{x \in A}$ on a finite or countable support, with $p_x > 0$. A sampler has a finite reachable state set S, a fixed initial state s_0, and a time-homogeneous deterministic rule. At each state it reads a word c from a prefix-free binary stopping code and emits an output x and a next state t. The code is exhaustive: its stopping cylinders have total fair-bit probability one. Codes may be infinite, and the empty word is allowed. Fresh unused fair bits are used at each round. The emitted process is required to be exactly iid with law p from s_0.

Only the state retained between outputs is counted. There is no free external iteration counter. The abstract lower-bound model does not require a finite representation of the target or a uniform computable procedure for all targets. Constructive upper bounds use computable CDF access; every fixed finite algebraic law admits such access. Temporary within-round work is unrestricted.

Write $I_x = \log(1/p_x)$, let ℓ_i be the consumed length in round i, and set

$$T_n = \sum_{i=1}^{n} \ell_i, \qquad L_n = \sum_{i=1}^{n} (\ell_i - I_{X_i}).$$

Each realized summand is finite. For an actual finite path of total input length L and output word w, its input cylinder has probability 2^{-L} and is contained in the event of producing w. Exactness gives

$$2^{-L} \leq \prod_{x \text{ in } w} p_x, \qquad L_n \geq 0. \tag{1}$$

Prefix-freeness makes this containment valid even with empty stopping words. Thus $\mathbb{E}L_n \in [0, \infty]$ is well defined without splitting expectations.

Definition 2.1. A sampler has zero asymptotic expected path loss if $\mathbb{E}L_n = o(n)$. If $H(p) < \infty$, this is equivalent to $\mathbb{E}T_n = nH(p) + o(n)$. At infinite entropy we use only the path-loss formulation and never subtract infinite expectations.

The following potential is the state-information mechanism of [1]; we include its short construction for the exact model used here.

Lemma 2.2 (Nonnegative defects)**.** *There is a finite nonnegative function $h : S \to \mathbb{R}$, with $h(s_0) = 0$, such that every actual transition (s, c, x, t) has*

$$d(s, c, x, t) := |c| - I_x + h(s) - h(t) \geq 0.$$

Moreover,

$$L_n = \sum_{i=1}^{n} d_i + h(S_n). \tag{2}$$

Proof. Define $h(s)$ as the infimum of $L - \sum I_x$ over finite actual paths from s_0 to s. Equation (1) makes it nonnegative. Reachability gives a finite witnessing path, so it is finite; the empty path gives $h(s_0) = 0$. Append a transition to paths approaching this infimum to obtain $h(t) \le h(s) + |c| - I_x$. Telescoping proves (2). The function is bounded because S is finite. □

3 The countable zero-loss classification

Theorem 3.1 (Countable rigidity)**.** *A finite or countable discrete law admits an exact iid sampler in the model of Section 2 with finitely many persistent states and $\mathbb{E}L_n = o(n)$ if and only if it is a reciprocal-power law. In the positive case a one-state sampler has $L_n = 0$ on every stopping path. When $H(p) < \infty$, it also has $\mathbb{E}T_n = nH(p)$.*

We first record the finite algebraic normalization needed in the proof.

Lemma 3.2 (Positive rational normalization)**.** *Suppose $J < \infty$, $b_j > 0$ are rational, $r_j \ge 0$ are rational, and $\sum_{j=1}^{J} b_j 2^{-r_j} = 1$. Then every r_j is an integer.*

Proof. Choose a common denominator q and write $r_j = n_j/q$. For $\alpha = 2^{1/q}$, Eisenstein irreducibility of $X^q - 2$ gives independence over $\mathbb{Q}$ of $1, \alpha^{-1}, \dots, \alpha^{-(q-1)}$. Grouping n_j by residues modulo q expresses the normalization in this basis with nonnegative rational coefficients. Every nonzero residue must have coefficient zero. Positivity forces each n_j to be divisible by q. □

Proof of Theorem 3.1. Assume first $\mathbb{E}L_n = o(n)$ and use Lemma 2.2.

Recurrent transitions are tight. Let $g(s)$ be the expected one-step defect conditional on current state s. The nonnegative cumulative defect sums increase with n, and their expectations are finite for all n since $\mathbb{E}L_n$ is finite eventually. Positive-probability reachability then implies $g(s) < \infty$ at every reachable state. Let ν_i be the state law after i outputs. Equation (2) gives

$$\frac{1}{n} \sum_{i=0}^{n-1} \nu_i g \longrightarrow 0.$$

The finite state chain has a Cesaro limiting stationary law π, positive on precisely the union K of its reachable closed recurrent classes. Each such class is reached with positive probability through a finite witnessing path. Hence $\pi g = 0$ implies $g(s) = 0$ for all $s \in K$. Every individual stopping word has positive probability; all transitions out of K therefore have $d = 0$. This union is closed. The finite-state Markov property follows from independence of the unused fair-bit suffix after an almost-surely finite prefix stopping time; finite expected stopping length is unnecessary.

Stationary output windows remain iid. For each x let $P_x(s,t)$ be the one-round probability of output x and next state t. For $w = x_1 \cdots x_k$ put $P_w = P_{x_1} \cdots P_{x_k}$ and $p_w = \prod_j p_{x_j}$. Exactness gives $\nu_i P_w \mathbf{1} = p_w$ for every i. Cesaro averaging yields

$$\pi P_w \mathbf{1} = p_w. \tag{3}$$

Restrict to the closed union K.

Only finitely many integer kernel types occur. Tightness fixes the length of every stopping word for a given triple (s, x, t):

$$|c| = I_x + h(t) - h(s).$$

Thus its multiplicity $N_x(s,t)$ is a finite nonnegative integer. With $D = \mathrm{diag}(2^{h(s)})$ on K,

$$P_x = p_x D N_x D^{-1}. \tag{4}$$

Set $u = \pi D$ and $v = D^{-1}\mathbf{1}$. Both vectors are strictly positive, $uv = 1$, and the one-letter case of (3) gives $uN_x v = 1$. Every entry is consequently bounded by $1/(u_s v_t)$, uniformly in x. Finite dimension and integrality imply that only finitely many distinct matrices N_x occur. No semigroup finiteness theorem is needed.

Each information value is rational. Put $m = |K|$. Since $\pi P_x^m \mathbf{1} = p_x^m > 0$, there is an actual m-step path within K emitting only x. It has a cycle of some length $1 \le k \le m$. Summing tight defects around that cycle gives kI_x equal to its total integer input length. Hence I_x is rational.

A finite exact rational normalization. Choose representatives $r_1, \dots, r_J$ of the integer types; write $N_j = N_{r_j}$ and $C_j = P_{r_j}$. Each C_j has dyadic-rational entries, because each entry counts finitely many stopping words of a single length. If x has type j, any common nonzero entry gives $I_x - I_{r_j} \in \mathbb{Z}$, and (4) gives $P_x = (p_x/p_{r_j})C_j$. The numbers

$$a_j = \sum_{x:\, N_x = N_j} p_x/p_{r_j}$$

are finite and strictly positive. Nonnegative summation gives $P = \sum_j a_j C_j$, which is stochastic. The rational affine system

$$\sum_j b_j C_j \mathbf{1} = \mathbf{1} \tag{5}$$

therefore has a strictly positive real solution. It has a strictly positive rational solution as well: rational Gaussian elimination provides a rational particular solution and rational nullspace basis, and rational free coordinates can approximate those of (a_j) closely enough to preserve strict positivity. Equation (5) remains exact.

Conjugating (5) gives $Qv = v$ for $Q = \sum_j b_j p_{r_j} N_j$. Applying u yields

$$1 = uQv = \sum_j b_j p_{r_j},$$

because $uv = uN_j v = 1$. Lemma 3.2 now applies to this *finite* positive rational normalization: every I_{r_j} is integral. The integer same-type differences prove that every I_x is integral. No infinite radical grouping has been used.

Conversely, if $p_x = 2^{-\ell_x}$ with integral $\ell_x \ge 0$, countable Kraft equality supplies a prefix code with these lengths. Its stopping cylinders have total probability one; nonstopping rays form a null set. Return to the same state after each output. Fresh suffix independence gives iid p, and every realized length is exactly I_x. All countably many rounds stop almost surely. Thus $L_n = 0$, including at infinite entropy. Finite entropy gives the additional expected-bit identity. The singleton mass-one law uses the empty word. □

4 Signed charges and finite-block bounds

For this section A is finite. Index its r positive masses and write $I = (I_x)$. Let $a = (a_x) \in \mathbb{Z}^r \setminus \{0\}$ satisfy

$$c \cdot I \in \mathbb{Z} \implies c \cdot a = 0 \qquad (c \in \mathbb{Z}^r). \tag{6}$$

Put $A_a = \max_x |a_x|$. For $M \ge 1$ define

$$\rho_a(M) = \min_{\substack{c \in \mathbb{Z}^r_{\ge 0},\ 1 \le \sum_x c_x \le M \\ c \cdot a \ne 0}} \frac{\lceil c \cdot I \rceil - c \cdot I}{|c \cdot a|}. \tag{7}$$

The admissible set is nonempty, since some single symbol has nonzero charge. It is finite, and (6) makes every numerator strictly positive.

Theorem 4.1 (Signed block charge)**.** *Let an exact iid sampler for finite p have at most M persistent states. Let $Y_i = a_{X_i}$ be iid under p. For all integers $k \geq 1, n \geq 0$,*

$$\mathbb{E}T_n - nH(p) \geq \left\lfloor \frac{n}{k} \right\rfloor \rho_a(M) \left(\mathbb{E} \left| \sum_{i=1}^{k} Y_i \right| - A_a(M-1) \right). \tag{8}$$

The right side can be replaced by its maximum with zero.

Proof. Erase simple directed cycles from any realized k-step state walk until a simple residual path remains. Erasure partitions the actual edges into cycles of length at most M and a residual of length at most $M-1$. Edges joined by an erasure have matching endpoint states, so this partition preserves actual labels, input lengths and defects.

A cycle with count vector c has integer total input length L and total defect $L - c\cdot I \geq \lceil c\cdot I \rceil - c\cdot I$. If $c \cdot a \neq 0$, it pays at least $\rho_a(M)|c \cdot a|$; otherwise its defect is nonnegative. The residual's absolute charge is at most $A_a(M-1)$. The triangle inequality gives the pathwise estimate

$$\sum_{\text{block}} d_i \geq \rho_a(M) \left(\left| \sum_{\text{block}} a_{X_i} \right| - A_a(M-1) \right).$$

Partition the first n steps into full k-blocks and a remainder. Remainder defects and the terminal potential are nonnegative. Every block has the unconditional iid output law, so taking expectations and using finite entropy proves (8). Infinite expected bit consumption makes the bound automatic. □

Lemma 4.2 (Existence of a charge)**.** *Every finite law outside the reciprocal-power class admits an integer vector satisfying* (6)*.*

Proof. Let $\Lambda = \{c \in \mathbb{Z}^r : c \cdot I \in \mathbb{Z}\}$. If its rank were r, r independent integer rows would give an invertible rational matrix C with CI integral. Hence all I_x would be rational. Lemma 3.2 with weights one and $\sum_x p_x = 1$ would make them integral, a contradiction. Thus the rational span of Λ has a nonzero integer orthogonal vector a. □

Proposition 4.3 (Drift and fluctuations)**.** *Let $\lambda_p = \liminf_{n\to\infty}(\mathbb{E}T_n - nH(p))/n$ for the sampler under consideration. If $\mu = \mathbb{E}Y_1 \neq 0$, then $\lambda_p \geq \rho_a(M)|\mu|$. If $\mu = 0$, put $\sigma^2 = \mathbb{E}Y_1^2 > 0$ and $C_4 = 3 + \mathbb{E}Y_1^4/\sigma^4$. Then*

$$\lambda_p \geq \frac{\rho_a(M)}{k} \left(\frac{\sigma\sqrt{k}}{\sqrt{C_4}} - A_a(M-1) \right) \quad (k \geq 1).$$

In particular this bound is positive for a suitable $k = O_p(M^2)$.

Proof. For nonzero drift, $\mathbb{E}|\sum_1^k Y_i| \geq k|\mu|$. The rate is at least each fixed-k consequence of (8); take the supremum over k. For zero drift, writing $S_k = \sum_1^k Y_i$ gives

$$\mathbb{E}S_k^2 = k\sigma^2, \qquad \mathbb{E}S_k^4 = k\mathbb{E}Y_1^4 + 3k(k-1)\sigma^4 \leq C_4 k^2 \sigma^4.$$

Hölder interpolation yields $\mathbb{E}|S_k| \geq (\mathbb{E}S_k^2)^{3/2}/(\mathbb{E}S_k^4)^{1/2}$. For example, take $k = \lceil 4A_a^2 C_4 M^2/\sigma^2 \rceil$. The bracket is then at least $A_a(M+1) > 0$. □

5 Algebraic targets and logarithmic memory

Theorem 5.1 (Polynomial loss obstruction)**.** *For every fixed finite law p with positive real algebraic masses outside the reciprocal-power class, there are constants $c_p, D_p > 0$ such that every exact iid sampler with at most M persistent states satisfies*

$$\liminf_{n\to\infty} \frac{\mathbb{E}T_n - nH(p)}{n} \geq c_p M^{-D_p} \qquad (M \geq 1).$$

Proof. Let $\Gamma = \langle 2, p_1, \ldots, p_r\rangle$ be the multiplicative group of positive real numbers generated by these elements. It is finitely generated and torsion free, hence isomorphic to $\mathbb{Z}^d$. Rank one would express $2 = g^b$ and $p_x = g^{b_x}$ with integers $b \neq 0, b_x$, making every $I_x = -b_x/b$ rational. Lemma 3.2 would then put p in the exceptional class. Thus $d \geq 2$. An integer linear functional annihilating the nonzero vector of 2 gives a nonzero character $\chi : \Gamma \to \mathbb{Z}$ with $\chi(2) = 0$. It is nonzero on at least one p_x, since the listed elements generate Γ. Put $a_x = \chi(p_x)$. This satisfies (6).

For every count vector in (7) set $s = c \cdot I$ and

$$R = 2^{\lceil s\rceil} \prod_x p_x^{c_x} = 2^{\lceil s\rceil - s}.$$

Nonzero charge gives $1 < R < 2$. These are products of the fixed positive algebraic bases $2, p_1, \ldots, p_r$ with integer exponents of absolute value at most $K_p M$ for a fixed $K_p \geq 1$.

We use the real-algebraic product form of Matveev's theorem [4]; a stated form is [5, Theorem 2]. Write $h_{\mathrm{abs}}(\gamma)$ for the absolute logarithmic height, normalized with natural logarithms. For fixed positive algebraic bases γ_j in a real number field of degree D, choose $A_j \geq \max\{Dh_{\mathrm{abs}}(\gamma_j), |\ln \gamma_j|, 0.16\}$. If $\prod_j \gamma_j^{b_j} \neq 1$ and $B \geq \max\{1, |b_j|\}$, the theorem gives

$$\left|\prod_j \gamma_j^{b_j} - 1\right| > \exp\left[-1.4\, 30^{t+3} t^{4.5} D^2 (1 + \ln D)(1 + \ln B) \prod_{j=1}^{t} A_j\right].$$

Zero exponents may be omitted; the finitely many active subsets permit uniform constants. Since our bases, field and heights are fixed, $R - 1 \geq cM^{-C}$ for positive constants depending only on p. As $\ln R \geq (R-1)/2$ for $1 < R < 2$, the ceiling gap is at least $c'M^{-C}$. Since $|c \cdot a| \leq A_a M$,

$$\rho_a(M) \geq c'' M^{-(C+1)}.$$

Nonzero drift in Proposition 4.3 gives the asserted polynomial bound directly. For zero drift use its explicit $k = O_p(M^2)$ and bracket at least $A_a(M+1)$ to obtain $c'''M^{-(C+2)}$. Decreasing a positive constant covers all $M \geq 1$ and unifies both cases. □

Remark 5.2. Constants can be evaluated once the integer character and algebraic field and height data are supplied. For rational p, an odd-prime valuation nonzero on some mass gives a convenient explicit character. Lemma 4.2 also applies to arbitrary real targets; the polynomial Diophantine estimates concern the stated algebraic class.

Let $b_p(\varepsilon)$ be the minimum of $\lceil \log |S| \rceil$ over exact iid samplers with $\limsup_n (\mathbb{E}T_n/n - H(p)) \leq \varepsilon$.

Corollary 5.3 (Sharp order in persistent bits)**.** *For a fixed finite algebraic law outside the reciprocal-power class,*

$$b_p(\varepsilon) = \Theta_p(\log(1/\varepsilon)) \quad \text{as } \varepsilon \downarrow 0.$$

For a reciprocal-power law, $b_p(\varepsilon) = 0$. For $0 < \varepsilon \leq 1/2$ and $L = \lceil \log(1/\varepsilon)\rceil$, the known upper sampler gives

$$b_p(\varepsilon) \leq 2L + 2\lceil \log(L+1)\rceil + 20.$$

Proof. Theorem 5.1 gives $M \geq (c_p/\varepsilon)^{1/D_p}$ for a sampler with the required lim sup bound, proving the logarithmic lower bound.

For the upper bound we specialize the existing sampler of [3, Algorithm 3 and Proposition 6.1], using its one-step estimate under the hypothesis $\Delta \geq 1000$ of its Theorem 5.6. With a constant integer cap satisfying that hypothesis it retains an integer pair (Z, R) with $0 \leq Z < R \leq \Delta$, initially $(0, 1)$, and satisfies

$$\mathbb{E}T_n - nH(p) \leq \log \Delta + 48n\frac{\log \Delta}{\Delta}. \tag{9}$$

Its conditional output guarantee gives exact iid p across repeated calls. Fixed target and cap make the rule homogeneous; no increasing counter is retained. Its CDF approximation interface is available for finite algebraic targets. The number of pairs is $\Delta(\Delta+1)/2$; temporary oracle and arithmetic work is not counted as persistent state.

Set $W = L + \lceil \log(L+1) \rceil + 10$ and $\Delta = 2^W$. Since $L \geq 1$, this cap is at least 4096, $W \leq 2L + 10 \leq 6(L+1)$, and $\Delta \geq 1024\, 2^L(L+1)$. Thus $48W/\Delta \leq (288/1024)2^{-L} \leq \varepsilon$. There are fewer than Δ^2 pairs, so $2W$ bits suffice. This proves the stated upper estimate and the order claim. The one-state prefix construction proves the exceptional case. □

Remark 5.4. The upper sampler and its leading upper coefficient 2 are existing results; see [1, Section 8]. The new part of Corollary 5.3 is the lower bound for every finite algebraic nonexceptional target, including cancelling resonances. Neither an optimal leading coefficient in bits nor a matching power-law exponent in M has been found.

6 A fully resonant example and the block-method optimum

Consider 48 outcomes: three with mass 3/16, eighteen with mass 1/48, and twenty-seven with mass 1/432. The group masses are $9/16, 6/16, 1/16$ and sum to one. With $a_x = v_3(p_x)$, the information is $I_x = 4 - a_x \log 3$, and the charge Y has values $1, -1, -3$ with probabilities $9/16, 6/16, 1/16$. Therefore

$$\mathbb{E}Y = 0, \qquad \mathbb{E}Y^2 = 3/2, \qquad \mathbb{E}Y^4 = 6, \qquad H(p) = 4.$$

Multiplying the probability vector by 432 gives positive integer counts $81, 9, 1$ on the respective atoms, total count 432 and total information 1728. Thus a zero-gap count vector contains every symbol. Any bound that minimizes a ceiling gap per occurrence of one marked symbol over all count vectors of size at most M vanishes for $M \geq 432$.

The signed bound remains positive. Here $\mathbb{E}S_k^4 = (27k^2 - 3k)/4$, so $\mathbb{E}|S_k| \geq \sqrt{k/2}$. Taking $k = 72(M-1)^2$ in Theorem 4.1 gives

$$\lambda_p \geq \frac{\rho_a(M)}{24(M-1)} > 0 \qquad (M \geq 2).$$

If $\delta_a(M)$ is the minimum unnormalized ceiling gap over nonzero-charge count vectors of size at most M, then $\rho_a(M) \geq \delta_a(M)/(3M)$. Taking $k = 72M^2$ instead gives the convenient all-M estimate $\lambda_p \geq \delta_a(M)/(72M^2)$.

Corollary 6.1 (A concrete resonant exponent)**.** *For this* 48*-outcome law, every* $\eta > 0$ *admits a constant* $c_\eta > 0$ *such that every sampler with at most* M *persistent states has*

$$\lambda_p \geq c_\eta M^{-(6.116201+\eta)}.$$

Consequently,

$$\liminf_{\varepsilon \downarrow 0} \frac{b_p(\varepsilon)}{\log(1/\varepsilon)} \geq \frac{1}{6.116201}.$$

Proof. Write $\alpha = \ln 3/\ln 2$ and $\kappa = 4.116201$. We use the established linear-form measure bound $\nu(1, \ln 2, \ln 3) \le \kappa$ of Bondareva–Luchin–Salikhov [6, Theorem 1], also stated in [1, Section 6.1, Definition 16 and equation (45)]. Its two-sided absolute-value formulation implies that, for every $\eta > 0$, nonzero integer linear forms in these fixed numbers, of sufficiently large coefficient height H, have absolute value at least $H^{-(\kappa+\eta)}$.

For a count vector of length $m \le M$ and nonzero charge $v = c \cdot a$, its information is $s = 4m - v\alpha$, where $|v| \le 3M$. With $z = \lceil -v\alpha \rceil$ its ceiling gap satisfies

$$\delta = \lceil s \rceil - s = z + v\alpha \in (0, 1), \qquad \delta \ln 2 = z \ln 2 + v \ln 3.$$

This form is nonzero and its integer coefficient height is at most KM, for a fixed $K > 3\alpha + 1$. The absolute-value estimate covers both signs of v. The finitely many coefficient pairs below its threshold have positive gaps and can be absorbed into a positive constant. Hence $\delta_a(M) \ge c'_\eta M^{-(\kappa+\eta)}$. Apply the all-$M$ bound above to obtain the loss estimate. For each fixed η, a sampler of loss rate at most ε must have

$$\lceil \log M \rceil \ge \frac{\log(1/\varepsilon) + \log c_\eta}{6.116201 + \eta}.$$

First let $\varepsilon \downarrow 0$ and then $\eta \downarrow 0$. □

Remark 6.2. The logarithmic-form exponent is prior work. This corollary applies it through the signed bound to a target on which every marked-symbol bound eventually vanishes. The known upper leading coefficient remains 2.

Proposition 6.3 (Optimum of the block expression)**.** *Let Y be bounded, centered and nondegenerate, with variance σ^2, and let S_k be an iid partial sum. Then*

$$F(b) := \sup_{k \in \mathbb{N},\ k \ge 1} \frac{\mathbb{E}|S_k| - b}{k} \sim \frac{\sigma^2}{2\pi b} \quad (b \to \infty).$$

Also $F(b) \le \sigma^2/(4b)$ for $b > 0$.

Proof. Cauchy–Schwarz gives $\mathbb{E}|S_k| \le \sigma\sqrt{k}$; maximizing $\sigma/t - b/t^2$ over $t > 0$ gives the finite-b upper bound. The classical iid CLT and uniform integrability give $\mathbb{E}|S_k|/\sqrt{k} \to q = \sigma\sqrt{2/\pi}$. Uniform integrability of these absolute first moments follows from their uniformly bounded second moments, since the tail expectation at threshold R is at most σ^2/R.

For $\eta > 0$, the expectation is bounded above by $(q+\eta)\sqrt{k}$ for all sufficiently large k. The finitely many smaller k have nonpositive numerators for sufficiently large b. Continuous optimization therefore gives $\limsup_{b\to\infty} bF(b) \le (q+\eta)^2/4$. For a lower bound choose an integer k nearest $(2b/q)^2$ and use the expectation limit. This gives $\liminf bF(b) \ge q^2/4$. Let $\eta \downarrow 0$. □

For the 48-outcome law the optimized expression in Theorem 4.1 is consequently asymptotic to $\rho_a(M)/(4\pi(M-1))$. This statement is sharp only for that particular block expression.

7 Countable conditional dyadic refinements

Let $q = (q_1, \ldots, q_r)$ be a fixed finite law with positive masses. For each j choose a finite or countable set K_j and nonnegative integers $\ell_{j,k}$ with

$$\sum_{k \in K_j} 2^{-\ell_{j,k}} = 1, \qquad p_{j,k} = q_j 2^{-\ell_{j,k}}. \tag{10}$$

The output labels are the distinct pairs (j,k). No entropy bound is imposed. Define $b_p^L(\varepsilon)$ by minimizing $\lceil \log |S| \rceil$ over exact samplers with $\limsup_n \mathbb{E}L_n/n \le \varepsilon$. At finite entropy this agrees with $b_p(\varepsilon)$.

Theorem 7.1 (Uniform refinement transfer)**.** *Suppose q is outside the reciprocal-power class. Let $a \in \mathbb{Z}^r \setminus \{0\}$ satisfy (6) for q, put $A_a = \max_j |a_j|$, and define $\rho_{q,a}(M)$ by (7) for q. Let J_i be iid with law q and put $Y_i = a_{J_i}$. Every exact sampler for (10) with at most M persistent states satisfies*

$$\mathbb{E}L_n \ge \left\lfloor \frac{n}{k} \right\rfloor \rho_{q,a}(M) \left(\mathbb{E} \left| \sum_{i=1}^{k} Y_i \right| - A_a(M-1) \right) \qquad (k \ge 1, n \ge 0). \tag{11}$$

If the q_j are algebraic, constants $c_q, D_q > 0$, independent of the refinement, give

$$\liminf_n \frac{\mathbb{E}L_n}{n} \ge c_q M^{-D_q}.$$

If q is algebraic and fixed computable complete conditional prefix codes of lengths $\ell_{j,k}$ are supplied, then

$$b_p^L(\varepsilon) = \Theta_q(\log(1/\varepsilon)).$$

Both bounding constants are uniform over these refinements. The explicit upper estimate of Corollary 5.3 also holds for b_p^L.

Proof. Use Lemma 2.2, which requires no entropy integrability, and assign charge a_j to (j,k). On an erased simple cycle, let c_j count the first-coordinate labels, and let B be the sum of their conditional lengths. The cycle information is

$$s_p = \sum_j c_j \log(1/q_j) + B.$$

Although B need not be bounded in terms of M, it is an integer. Its contribution cancels exactly from $\lceil s_p \rceil - s_p$ before any logarithmic-form estimate is applied. The cycle charge is $c \cdot a$, so its nonnegative defect pays at least $\rho_{q,a}(M)|c \cdot a|$ when the charge is nonzero. The residual path has at most $M-1$ edges and charge at most $A_a(M-1)$. The pathwise proof of Theorem 4.1 therefore applies unchanged. Unconditional iid refined outputs have iid first coordinates with law q by (10). Taking expectations gives (11), even when $\mathbb{E}L_n$ is infinite. Proposition 4.3 and the finite-q logarithmic-form argument in Theorem 5.1 give the polynomial rate with constants depending only on q.

For the upper bound, run the existing finite-q sampler of Corollary 5.3. After a macro output j and next retained state t have been selected, sample k using fresh bits and the supplied conditional prefix code, then emit (j,k) and retain t. Concatenated stopping codes remain prefix-free and exhaustive. Every round stops almost surely. For any complete macro output/state path, specifying the conditional leaves multiplies its probability by $\prod_i 2^{-\ell_{j_i,k_i}}$. Summing over macro state paths proves iid refined outputs; it does not require conditional iid macro outputs given a hidden state.

The conditional code uses only temporary work and adds exactly $\ell_{j,k}$ input bits. On every stopping path,

$$(\text{macro length} + \ell_{j,k}) - \log(1/p_{j,k}) = \text{macro length} - \log(1/q_j).$$

Thus the lifted L_n equals the macro loss pathwise, with no additional persistent states. Its expected loss can be finite even when target entropy and expected input length are infinite. The known macro upper bound and the transferred polynomial lower bound prove the memory order. □

For a concrete infinite-entropy example, take $q = (1/3, 2/3)$ and, for each j, use leaves (m, w) with $m \geq 1$, $w \in \{0,1\}^{2^m}$ and codeword $1^{m-1}0w$. This is a computable prefix code: the total conditional mass at level m is 2^{-m}, while its code length is $m + 2^m$. Hence total mass is one and conditional entropy is $\sum_{m\geq 1} 2^{-m}(m + 2^m) = \infty$. All masses $q_j 2^{-(m+2^m)}$ are rational, every round stops almost surely, and Theorem 7.1 applies in expected-path-loss language.

For refinements of the 48-outcome law, the same cycle-gap identity and centered charge law transfer Corollary 6.1 as well: the rate is at least $c_\eta M^{-(6.116201+\eta)}$ and the limiting lower coefficient for b_p^L is at least $1/6.116201$. These are lower transfers. Coarsening a sampler generally discards information and does not preserve its loss. The upper composition is classical conditional prefix coding combined with the existing macro sampler.

8 Clock conventions and open quantitative questions

Several ingredients have substantial predecessors. Kraft coding supplies the reciprocal-power construction [7]. Kirchhoffer's finite-phrase binary variable-to-variable coding theorem uses unary words and radical arithmetic [8, Section 3.3]; see also [9]. Its finite reset-phrase model differs from arbitrary retained state with infinite per-round codes. Our countable argument does not normalize an infinite algebraic sum: finite integer kernel types instead yield a finite positive rational normalization. Cycle erasure and cycle-ratio charging are standard graph ideas [10]; the potential and information-cycle framework are taken from [1]. The logarithmic-form input is the established theorem of Matveev [4, 5].

Information cocycles, cycle-weight groups and fluctuation obstructions also arise in finitary coding. Parry–Schmidt [11, Theorem 6.2] already use additive characters of multiplicative cycle groups and a central limit theorem to prove invariance of character means and variances under their finitary coding hypotheses. Schmidt [12, Theorems 4.1 and 5.1] proves beta-function invariance and rigidity of finite Bernoulli laws under finitary isomorphisms with finite expected forward and inverse coding lengths. See also Parry's survey [13] and Serafin's exposition [14, Section 2.2]. Harvey–Peres [15] compare information fluctuations with coding-boundary capacity for equal-entropy, shift-commuting finitary homomorphisms; Gabor [16] determines the optimal range of coding-radius moments for equal-entropy finitary isomorphisms. These maps commute with a common unit shift; their spatial coding radius is a different resource from retained sampler state. Arithmetic characters and character-fluctuation rigidity are established tools. The present increment is a nonnegative sampling-defect inequality controlled by persistent state cardinality, with a variable number of input bits per output, and the countable finite-type normalization. The classical example below makes the clock distinction explicit. Coding precision, phrase delay and persistent state are not identified without a resource-preserving reduction.

Krieger's earlier directional coding invariant [17] also allows countably infinite Markov alphabets and assumes finite expected forward coding time. Thus finite support or inverse integrability alone does not separate our problem from all classical coding results. The relevant distinction remains the clock and the quantitative retained-state resource.

The distinction from a common-clock finitary coding problem can be seen in the classical Meshalkin pair recalled by Harvey–Peres [15, Introduction]. Let $p = (1/2, 1/8, 1/8, 1/8, 1/8)$. The complete prefix code

$$\{0,\ 100,\ 101,\ 110,\ 111\}$$

gives a one-state exact iid sampler with zero path loss. The target has entropy 2 and informational variance 1, whereas a uniform four-symbol source obtained by grouping fair bits in fixed pairs has entropy 2 and variance 0. Theorem 1 of [15] rules out a common-clock shift-commuting finitary

homomorphism for this pair with a finite square-root coding-radius moment; it does not rule out the variable-bit-clock sampler just displayed. Lack of invertibility alone would not distinguish the models, since their theorem already concerns homomorphisms. The pair and prefix construction clarify the resource model.

The exact sampling endpoint is settled here on countable support. Quantitative bounds cover all finite algebraic targets and their conditional dyadic refinements. Open questions include optimal leading coefficients in $b_p(\varepsilon)$, matching loss exponents in the state count, general countable quantitative bounds, and the role of exceptionally well approximable transcendental probabilities.

Acknowledgment of tools. AI-assisted tools were used in exploring and internally checking the arguments and in preparing the manuscript.

References

[1] T. L. Draper and F. A. Saad, *Space–Entropy Lower Bounds for Random Sampling*, arXiv:2607.14503v1, 2026. https://arxiv.org/abs/2607.14503.

[2] T. L. Draper and F. A. Saad, *Efficient Online Random Sampling via Randomness Recycling*, Proceedings of SODA 2026, 2473–2511; arXiv:2505.18879v5. https://doi.org/10.1137/1.9781611978971.89.

[3] T. L. Draper, D. G. Harris and F. A. Saad, *Online Random Sampling with Real Probabilities*, arXiv:2607.13828v1, 2026. https://arxiv.org/abs/2607.13828.

[4] E. M. Matveev, *An explicit lower bound for a homogeneous rational linear form in the logarithms of algebraic numbers. II*, Izvestiya: Mathematics 64 (2000), 1217–1269. https://doi.org/10.1070/IM2000v064n06ABEH000314.

[5] G.-H. Jeong and P.-S. Park, *On the Diophantine equation* $F_n^2 + F_m^2 = 2^a$, Mathematical Communications 30 (2025), 171–177. https://doi.org/10.64785/mc.30.2.2.

[6] I. V. Bondareva, M. Yu. Luchin and V. Kh. Salikhov, *Symmetrized polynomials in a problem of estimating of the irrationality measure of number* $\ln 3$, Chebyshevskii Sbornik 19 (2018), no. 1, 15–25. https://doi.org/10.22405/2226-8383-2018-19-1-15-25.

[7] D. E. Knuth and A. C. Yao, *The complexity of nonuniform random number generation*, in Algorithms and Complexity: New Directions and Recent Results, Academic Press, 1976, 357–428.

[8] H. Kirchhoffer, *Design and application of variable-to-variable length codes*, doctoral dissertation, University of Rostock, 2016.

[9] H. Kirchhoffer, D. Marpe, H. Schwarz and T. Wiegand, *Properties and Design of Variable-to-Variable Length Codes*, ACM Transactions on Multimedia Computing, Communications, and Applications, 2018. https://doi.org/10.1145/3230653.

[10] M. Hartmann and J. B. Orlin, *Finding minimum cost to time ratio cycles with small integral transit times*, Networks 23 (1993), 567–574.

[11] W. Parry and K. Schmidt, *Natural coefficients and invariants for Markov-shifts*, Inventiones Mathematicae 76 (1984), 15–32. https://doi.org/10.1007/BF01388488.

[12] K. Schmidt, *Invariants for finitary isomorphisms with finite expected code lengths*, Inventiones Mathematicae 76 (1984), 33–40.

[13] W. Parry, *Notes on coding problems for finite state processes*, Bulletin of the London Mathematical Society 23 (1991), 1–33. https://doi.org/10.1112/blms/23.1.1.

[14] J. Serafin, *Finitary codes, a short survey*, in Dynamics and Stochastics, IMS Lecture Notes–Monograph Series 48 (2006), 262–273. https://doi.org/10.1214/074921706000000284.

[15] N. Harvey and Y. Peres, *An invariant of finitary codes with finite expected square root coding length*, arXiv:math/0309120, 2003. https://arxiv.org/abs/math/0309120.

[16] U. Gabor, *Moment-optimal finitary isomorphism for i.i.d. processes of equal entropy*, arXiv:2412.15425, 2024. https://arxiv.org/abs/2412.15425.

[17] W. Krieger, *On the finitary isomorphisms of Markov shifts that have finite expected coding time*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65 (1983), 323–328. https://doi.org/10.1007/BF00532486.