EXACTNESS, CRITICALITY, AND MEMORY IN MEAN-FIELD SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

How much information must a sequential generator retain after it has emitted part of a sample? We answer this question for rank-one quadratic Gibbs distributions in a write-only, finite-state interface with a free clock and read-only transition kernels. For every nonzero coupling and arbitrary finite external fields, the minimum state count at a cut is the smaller number of attainable weighted sums on its two sides. A single stochastic generator simultaneously attains these cutwise lower bounds by switching from a prefix sum to a sampled suffix sum. In the homogeneous Curie–Weiss model, exact generation consequently needs $\lfloor n/2 \rfloor + 1$ states at every nonzero coupling. Allowing fixed total-variation error produces a different classification: the optimal width is $\Theta_\varepsilon(n^{1/4})$ at criticality and $O_{\beta,\varepsilon}(1)$ at each fixed positive noncritical coupling. The critical lower bound uses a robust near-diagonal mixture inequality, while the upper bound quantizes a Gaussian auxiliary field. Finally, arbitrarily small dyadic heterogeneity can raise exact memory to $\lfloor n/2 \rfloor$ bits without changing the fixed-error classification. These results separate algebraic distinguishability from statistically necessary memory; they do not bound uniform implementation time or arithmetic workspace.

1 INTRODUCTION

A sequential generator must preserve enough information about its past outputs to produce the correct future dependence. It need not preserve the past itself. This distinction makes retained memory a distributional resource: the relevant question is not how many parameters describe a target, but how many internal configurations are necessary to connect its past and future.

Mean-field spin models expose an ambiguity in that question. Their interaction can be specified by one scalar statistic, yet this statistic can take polynomially or exponentially many distinct values. In addition, exact distinguishability can persist between distributions that are arbitrarily close in total variation. A lower bound based only on exact rank may therefore measure the insistence on exactness rather than a robust statistical obstacle.

We study these issues in an explicit finite-state interface. A generator emits spins in a prescribed order and never reads its output tape. Between emissions it retains a state from a finite alphabet. The clock and read-only transition kernels are not charged. All persistent randomness is charged. This interface allows arbitrary stochastic computation within a transition, so our upper bounds concern retained information rather than total implementation space. The lower bounds apply even under this generous computational access.

Our results reveal two different classifications. Exact state complexity is governed by attainable weighted sums and is insensitive to the magnitude or sign of a nonzero quadratic coupling. Fixed-error complexity in the homogeneous ferromagnet instead distinguishes the critical point from both neighboring phases. At criticality, a growing continuum of collective fluctuations must cross the middle of the generated sequence. Away from criticality, a bounded number of discretized auxiliary-field values suffices, including in the ordered phase.

Contributions. First, we characterize the exact state count at every cut and construct one generator attaining all of these bounds, not just the maximum lower bound at an isolated cut. Second, we prove a fixed-total-variation critical width law, $\Theta_\varepsilon(n^{1/4})$, using a support-cardinality converse that remains

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

valid for nonsymmetric approximating laws. Third, we show that vanishing dyadic perturbations can produce maximal exact width while preserving the approximate law, and quantify the role of input precision. Complete proofs, including the concentration estimates needed for the approximate upper bound, appear in the appendices.

Relation to prior work. Finite-state descriptions already provide a rigorous language for distributional expressivity of recurrent models (Svete and Cotterell, 2023). Our states are discrete retained configurations, not real-valued neurons. Nonnegative rank and one-shot generation of bipartite distributions are classical correlation-complexity tools (Jain et al., 2013). Ordinary cut ranks likewise underlie tensor-train representations (Oseledets, 2011). We use this lower-bound machinery rather than claim it as new. What must additionally be shown here is a *nonnegative, normalized, sequential realization attaining every cut bound* for the specified Gibbs family.

The critical magnetization scale and the logarithmic divergence of shared information are also established phenomena (Chatterjee and Shao, 2011; Wilms et al., 2012; Cohen et al., 2015). In particular, the $1/4$ coefficient for critical mutual information is not a new discovery of this paper. Our result concerns the minimum *support cardinality* of a sequential generator at a fixed nonzero distributional error. Mutual information of the target alone does not establish that operational statement for every nearby law. The auxiliary-field identity is classical (Hubbard, 1959); related transforms support efficient low-rank Ising sampling (Koehler et al., 2022). We study a different resource from the running time of those algorithms.

2 Interface and main results

For $n \geq 1$, finite $\beta \in \mathbb{R}$, weights $w \in \mathbb{R}^n$, and fields $h \in \mathbb{R}^n$, let

$$P_{\beta,w,h}(x) = Z^{-1} \exp\left\{ \frac{\beta}{2n} \left(\sum_{i=1}^{n} w_i x_i \right)^2 + \sum_{i=1}^{n} h_i x_i \right\}, \qquad x \in \{-1,+1\}^n. \tag{1}$$

The homogeneous zero-field law is $P_{n,\beta} = P_{\beta,\mathbf{1},0}$. Its ferromagnetic critical parameter is $\beta = 1$. All finite targets in (1) have full support. State-count results use base-two logarithms; entropy calculations use natural logarithms.

Definition 1 (Retained-state generator)**.** A generator consists of finite state alphabets $\mathcal{S}_0, \ldots, \mathcal{S}_n$, a deterministic initial state, and kernels $T_t(x, s' \mid s)$ for $t = 1, \ldots, n$. At step t it samples (X_t, S_t) from $T_t(\cdot, \cdot \mid S_{t-1})$ using fresh independent randomness and irrevocably emits X_t. The clock and the prescribed kernels are read-only and free. No other information, including a shared seed or an output-tape read, is available to future steps. The terminal state may be collapsed to one value.

Write Q_T for the output law. Its width is $\max_t |\operatorname{supp}(S_t)|$. Define

$$K_\varepsilon(P) = \min_{T:\mathrm{TV}(P,Q_T)\leq\varepsilon} \max_t |\operatorname{supp}(S_t)|, \qquad B_\varepsilon(P) = \lceil \log_2 K_\varepsilon(P) \rceil, \tag{2}$$

where $\mathrm{TV}(P,Q) = \frac{1}{2}\sum_x |P(x) - Q(x)|$. Thus B_ε counts bits needed to label retained states, *not* bits needed to store the transition table or evaluate its probabilities. Different layers may reuse the same labels. The transition at which the representation changes is determined by the free clock.

At cut t, define the distinct signed-sum sets

$$\begin{aligned} \mathcal{A}_t &= \left\{ \sum_{i\leq t} w_i x_i : x_i \in \{-1,+1\} \right\}, \quad a_t = |\mathcal{A}_t|, \\ \mathcal{C}_t &= \left\{ \sum_{i>t} w_i x_i : x_i \in \{-1,+1\} \right\}, \quad c_t = |\mathcal{C}_t|. \end{aligned} \tag{3}$$

Empty sums are zero, so $a_0 = c_n = 1$.

Theorem 2 (Exact cutwise complexity)**.** *For every $\beta \neq 0$ and finite w, h,*

$$K_0(P_{\beta,w,h}) = \max_{0\leq t\leq n} \min(a_t, c_t). \tag{4}$$

Every exact generator has $|\operatorname{supp}(S_t)| \geq \min(a_t, c_t)$ at every cut. One generator attains equality simultaneously at all cuts. If $\beta = 0$, the law is a product distribution and $K_0 = 1$.

Table 1: Homogeneous zero-field model. Approximate entries hold at fixed $0 < \varepsilon \le 1/8$; the critical lower bound is stated for even n. The approximation bounds concern full-vector total variation.

Coupling	Exact width K_0	Fixed-error width K_ε
$\beta = 0$	1	1
$0 < \beta < 1$	$\lfloor n/2 \rfloor + 1$	$\Theta_{\beta,\varepsilon}(1)$
$\beta = 1$	$\lfloor n/2 \rfloor + 1$	$\Theta_{\varepsilon}(n^{1/4})$
$\beta > 1$ fixed	$\lfloor n/2 \rfloor + 1$	$\Theta_{\beta,\varepsilon}(1)$

Theorem 3 (Fixed-error critical memory)**.** *There are absolute constants $c, C > 0$ such that, for every even $n \ge 2$ and $0 < \varepsilon \le 1/8$,*

$$cn^{1/4} \le K_\varepsilon(P_{n,1}) \le C\varepsilon^{-3/2} n^{1/4}. \tag{5}$$

For every fixed $\beta > 0$ with $\beta \ne 1$, there is $C_\beta < \infty$ such that, for all $n \ge 2$ and $0 < \varepsilon \le 1/8$,

$$1 \le K_\varepsilon(P_{n,\beta}) \le C_\beta \varepsilon^{-3/2}. \tag{6}$$

In particular, at fixed error $B_\varepsilon(P_{n,1}) = \frac{1}{4}\log_2 n + O_\varepsilon(1)$, whereas $B_\varepsilon(P_{n,\beta}) = O_{\beta,\varepsilon}(1)$ for each fixed positive noncritical β.

The constants are not asserted to remain bounded as $\beta \to 1$, and the dependence on ε is not claimed optimal. The even-n restriction concerns only the critical lower bound as stated. Table 1 contrasts exact and fixed-error memory.

3 Exact generation: a cut rank and a representation switch

At any cut of a generator, the past and future are conditionally independent given the retained state. Consequently the cut probability matrix factors through that state alphabet:

$$Q(u, v) = \sum_s Q(X_{\le t} = u, S_t = s)\, Q(X_{>t} = v \mid S_t = s). \tag{7}$$

For (1), deleting positive row and column factors leaves the matrix

$$E_t(a, c) = \exp(\beta ac/n), \qquad a \in \mathcal{A}_t, \quad c \in \mathcal{C}_t. \tag{8}$$

Distinct real exponentials form a Chebyshev system: a nonzero linear combination of r exponentials with distinct exponents has at most $r - 1$ distinct zeros. An induction using Rolle's theorem proves this directly. Thus (8) has rank $\min(a_t, c_t)$, giving the lower bound in Theorem 2. This is an exact algebraic statement, not a floating-point rank computation.

The matching generator uses two different sufficient statistics. Prefix support size a_t is nondecreasing, and suffix support size c_t is nonincreasing. Let $t_* = \max\{t : a_t \le c_t\}$. Until layer t_* retain the prefix sum. On the next transition, jointly sample the next emitted spin and the total weighted sum of the remaining suffix, conditional on the retained prefix. Thereafter retain only that remaining sum and decrement it as spins are emitted.

To make this constructive, set

$$G_t(c) = \sum_{x_{>t}:\, \sum_{i>t} w_i x_i = c} e^{\sum_{i>t} h_i x_i}, \qquad F_t(a) = \sum_{c \in \mathcal{C}_t} G_t(c) e^{\beta(a+c)^2/(2n)}. \tag{9}$$

Before the switch, the transition from prefix sum a emits x with probability

$$e^{h_{t+1}x} F_{t+1}(a + w_{t+1}x)/F_t(a). \tag{10}$$

At the switch, the joint probability of emitted spin x and new suffix state c is

$$e^{h_{t+1}x} G_{t+1}(c) e^{\beta(a+w_{t+1}x+c)^2/(2n)}/F_t(a). \tag{11}$$

Afterward, a suffix state c emits x with probability

$$e^{h_{t+1}x} G_{t+1}(c - w_{t+1}x)/G_t(c), \tag{12}$$

and retains $c - w_{t+1}x$. Unattainable sums have message zero. Once the suffix total has been sampled, the quadratic energy is constant over its possible completions; this is why the prefix can be discarded. Appendix A verifies normalization, the full output law, and all support counts, including degenerate weights.

For homogeneous weights, $a_t = t + 1$ and $c_t = n - t + 1$. This proves the exact column of Table 1. Conversely, if every subset sum is distinct, then $a_t = 2^t$ and $c_t = 2^{n-t}$, giving $B_0 = \lfloor n/2 \rfloor$. Rank-one interaction alone therefore does not control exact retained memory.

4 Approximation: a robust critical obstruction

The exact rank argument cannot by itself prove a fixed-error lower bound: nonzero singular values can be arbitrarily small. We instead use an inequality for a distribution concentrated near a spread-out diagonal.

Lemma 4 (Near-diagonal mixture bound)**.** *Suppose (U, V) take values in the same finite alphabet under P, with $P(U \neq V) \leq \delta$ and $\max_i P(U = i) \leq \alpha$. If $\mathrm{TV}(P, Q) \leq \varepsilon$ and Q is a mixture of at most K product distributions, then*

$$K\alpha \geq 1 - 2\delta - 2\varepsilon. \tag{13}$$

Proof. For one product distribution $a \otimes b$, select an index i maximizing a_i. Its diagonal mass outside (i, i) is at most its off-diagonal mass. Indeed, if $a_i \leq 1/2$ the entire diagonal mass is at most $1/2$; otherwise the omitted diagonal mass is at most $(1 - a_i)(1 - b_i) \leq a_i(1 - b_i)$. For a K-component mixture, let T contain the selected indices. Then $|T| \leq K$ and $Q(U = V \notin T) \leq Q(U \neq V)$. Hence $1 - \delta - K\alpha \leq P(U = V \notin T) \leq \delta + 2\varepsilon$. □

Critical lower bound. Let $n = 2m$, $A = \sum_{i \leq m} X_i$, $B = \sum_{i > m} X_i$, and $M = A + B$. Conditional on M, exchangeability gives

$$\mathbb{E}[(A - B)^2 \mid M] = \frac{n^2 - M^2}{n - 1} \leq 2n. \tag{14}$$

At $\beta = 1$, a direct Stirling estimate gives $\max_k P((M + n)/2 = k) \leq Cn^{-3/4}$. A hypergeometric marginalization transfers this bound to half-counts, up to a factor smaller than two. Consequently every interval of width $W\sqrt{n}$ contains at most $C(W + 1)n^{-1/4}$ marginal mass of A.

Quantize A and B into intervals of width $W\sqrt{n}$ using the same offset. Averaging the offset over one interval shows that some offset has disagreement probability at most $\mathbb{E}|A - B|/(W\sqrt{n}) \leq \sqrt{2}/W$. Take $W = 16$. For any approximating generator, the quantized variables form a mixture of at most $|\operatorname{supp}(S_m)|$ product distributions by (7); total variation cannot increase under quantization. Lemma 4 yields $|\operatorname{supp}(S_m)| \geq cn^{1/4}$ for $\varepsilon \leq 1/8$. No exchangeability of the approximating law is assumed. All finite-n estimates are proved in Appendix B.

A full-sequence upper bound. For $\beta > 0$, the Gaussian auxiliary-field identity gives

$$P_{n,\beta} = \int P_z \, \pi_{n,\beta}(dz), \qquad P_z(x) = \prod_{i=1}^{n} \frac{e^{\sqrt{\beta/n} z x_i}}{2\cosh(\sqrt{\beta/n} z)}, \tag{15}$$

$$\pi_{n,\beta}(dz) \propto e^{-z^2/2} \cosh(\sqrt{\beta/n} z)^n \, dz.$$

The log-partition curvature of P_z is at most β, so

$$D_{\mathrm{KL}}(P_z \| P_{z'}) \leq \tfrac{\beta}{2}(z - z')^2, \qquad \mathrm{TV}(P_z, P_{z'}) \leq \tfrac{\sqrt{\beta}}{2}|z - z'|. \tag{16}$$

This metric has a dimension-independent resolution. Truncate the latent distribution to mass at least $1 - \varepsilon/2$, then round it on a grid of mesh at most $\varepsilon/\sqrt{\beta}$. The resulting full-vector law is within ε of the target. A generator can sample the grid index jointly with its first output, retain only that index, and emit all subsequent coordinates independently. Thus the grid support bounds the width at every layer, not only at the middle cut.

It remains to control the length of the region that must be covered. Below criticality, $\mathbb{E}_\pi z^2 \le C_\beta$. Above criticality, the density has two wells centered at $\pm\sqrt{n}\,t_\beta$, where $t_\beta = \sqrt{\beta}\tanh(\sqrt{\beta}t_\beta) > 0$, and $\mathbb{E}_\pi \operatorname{dist}(z, \{\pm\sqrt{n}\,t_\beta\})^2 \le C_\beta$. The centers move apart, but their locations reside in the read-only kernels: only a well label and a discretized local fluctuation need be retained.

At criticality the quadratic term cancels. Writing $f(t) = t^2/2 - \log\cosh t$, one has $c\min(t^4, t^2) \le f(t) \le t^4/12$. These inequalities imply $\mathbb{E}_\pi z^2 \le C\sqrt{n}$. Chebyshev's inequality therefore supplies one or two intervals of total length $O_\beta(\varepsilon^{-1/2})$ away from criticality, and $O(\varepsilon^{-1/2}n^{1/4})$ at criticality. Grid covering proves the upper bounds of Theorem 3. Appendix C proves these bounds without exchanging an asymptotic limit and a discretization.

5 Exactness is unstable; finite precision limits the instability

Theorem 5 (Vanishing heterogeneity, maximal exact memory)**.** *Fix $n \ge 2$, $\beta \ne 0$, finite h, and $0 < \delta < 1$. With $w_i^{(\delta)} = 1 + \delta 2^{-i}$,*

$$\begin{aligned} B_0(P_{\beta,w^{(\delta)},h}) &= \lfloor n/2 \rfloor, \\ \mathrm{TV}(P_{\beta,w^{(\delta)},h}, P_{\beta,\mathbf{1},h}) &\le \min\left\{1, |\beta|\left(\delta + \frac{\delta^2}{2n}\right)\right\}. \end{aligned} \tag{17}$$

The exact-memory conclusion holds for every fixed coordinate permutation.

The binary fractional perturbation distinguishes equal-cardinality subsets; unequal cardinalities cannot cancel because the perturbation sum is less than one. Thus all subset sums are distinct. On the other hand, the perturbation changes every total weighted sum by less than δ. Its Hamiltonian change is uniformly at most the bound in (17). Interpolation of Gibbs laws converts this energy bound into a TV bound. Appendix D supplies both arguments.

If $\mathrm{TV}(P, P') \le r < \varepsilon$, the triangle inequality gives

$$K_{\varepsilon+r}(P) \le K_\varepsilon(P') \le K_{\varepsilon-r}(P). \tag{18}$$

Hence, for $h = 0$, fixed $\beta > 0$, $0 < \varepsilon \le 1/16$, and the bound in (17) at most $\varepsilon/2$, the fixed-error classification of Theorem 3 survives. Maximal exact memory is compatible with constant approximate memory away from criticality and $\Theta_\varepsilon(n^{1/4})$ approximate width at criticality.

Proposition 6 (Dyadic precision budget)**.** *Suppose $n \ge 2$, $\beta \ne 0$, $|w_i| \le 2$, and $w_i \in 2^{-b}\mathbb{Z}$, where $b \ge 0$ is an integer. Then*

$$B_0(P_{\beta,w,h}) \le \min\left\{\lfloor n/2 \rfloor, \left\lceil \log_2(2n2^b + 1)\right\rceil\right\}. \tag{19}$$

For even $n = 2m$, let $r = \min(b, m)$. There are such weights, for every finite h, with

$$K_0 = (m - r + 1)2^r. \tag{20}$$

In particular the worst-case exact bit complexity is $\Theta(\min\{n, b + \log_2 n\})$ over these dyadic weights, for even $n \ge 2$.

The construction mirrors a half-list consisting of $1 + 2^{-1}, \ldots, 1 + 2^{-r}$ and $m - r$ copies of 1. The fractional part identifies selected special weights, while the remaining integer count supplies $m - r + 1$ possibilities. The result also explains the precision cost of Theorem 5: choosing $\delta = 2^{-q}$ uses up to $n + q$ fractional bits, not a fixed-resolution alphabet.

6 Finite numerical diagnostics

All experiments run on a CPU, without model training or an external dataset. Exact combinatorial sum labels use rational arithmetic; probabilities use binary64. The reported errors are finite numerical diagnostics, not interval certificates or substitutes for the proofs. Appendix F specifies every parameter and the supplementary code reproduces every table and figure.

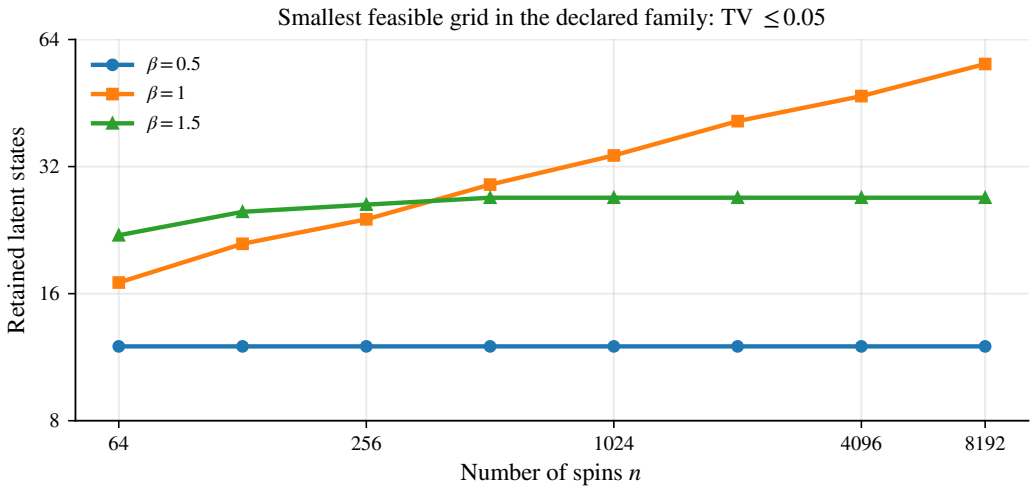

Figure 1: Widths of an explicit discretized-latent family at TV tolerance 0.05. Mesh selection is over $\{0.25, 0.5, 1, 2, 4, 8, 16, 32\}$. These are constructive upper diagnostics, not measured optimal widths. The $n^{1/4}$ behavior is established by Theorem 3, not inferred from this plot.

Checking the sequential realization. We enumerate both the actual state machine and the direct Gibbs law for 150 cases: $n = 1, \ldots, 10$, three weight families, and $\beta \in \{-1.5, 0, 0.5, 1, 2\}$, with seeded nonzero fields. All reachable width profiles agree with Theorem 2. The largest full-law TV discrepancy is 1.04×10^{-15} after upward rounding; the largest local normalization residual is 2.89×10^{-15}. Two 4×4 exponential-kernel witnesses have nonzero determinants computed in exact rational arithmetic. Neither test is used to infer full rank in growing dimension.

Approximate generators at fixed error. For $\beta \in \{0.5, 1, 1.5\}$ and $n \in \{64, 128, \ldots, 8192\}$, we discretize (15). Both target and approximation are exchangeable, so full-vector TV equals TV between their explicitly evaluated count distributions. Figure 1 uses the smallest feasible width among eight declared mesh sizes at error tolerance 0.05; it does not optimize over all generators. At $n = 8192$, the selected widths are respectively 12, 58, and 26, with TVs 0.01745, 0.00412, and 0.01794. The critical width grows while the two noncritical widths stabilize.

A separate fixed-mesh check at mesh 0.25 evaluates all 24 cases. Its largest TV is 3.12×10^{-4}; changing quadrature order from 8 to 16 changes the computed count law by at most 1.72×10^{-14}. These agreement checks do not certify the quadrature tail error. We also evaluate the coarse-bin lower-bound expression at six finite sizes. It is quantitatively loose, ranging from 1.10 to 1.51 on this grid; we report that limitation rather than interpret it as an empirical sharp converse.

Perturbation without a robust memory jump. At $n = 10$, $\beta = 1.5$, and $\delta \in \{2^{-2}, 2^{-4}, \ldots, 2^{-16}\}$, exact support counting gives width 32 for every nonzero perturbation, compared with width 6 at zero. Yet the directly enumerated TV decreases to 3.82×10^{-6} at the smallest amplitude, below the analytical energy bound throughout (Figure 2). The experiment illustrates why numerical rank thresholds cannot establish the exact-memory theorem.

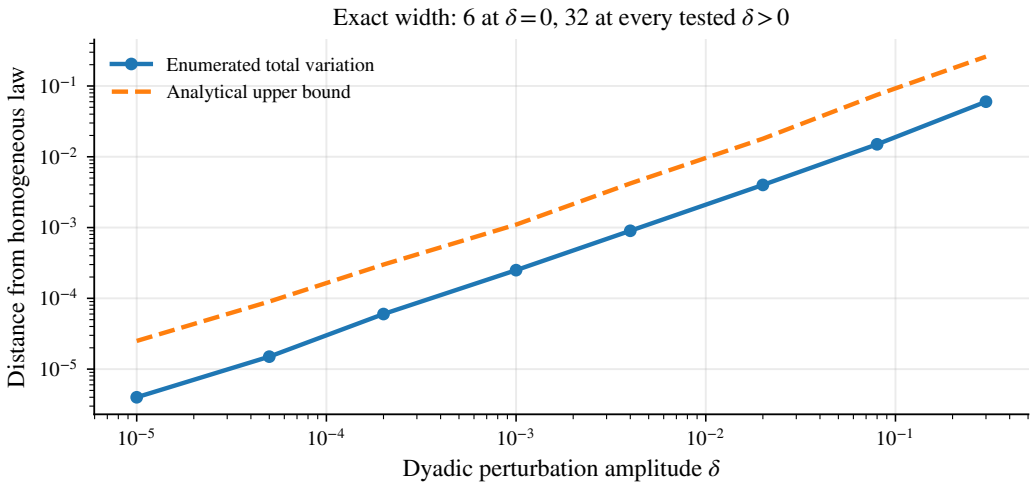

Figure 2: Arbitrarily close distributions can require different exact state counts. State counts are computed from rational sum supports, not from a singular-value threshold. The distance bound is Theorem 5; points are full-law enumerations.

7 Discussion and Limitations

The example separates three resources that are easily conflated: the number of distinct exact sufficient-statistic values, the number of statistically resolvable collective fluctuations, and the numerical cost of implementing a transition. Exact rank detects the first, while fixed-error retained memory detects the second. Our model intentionally does not measure the third.

The conclusions are restricted to rank-one interactions and a prescribed coordinate order. The approximate theorem is homogeneous and zero-field; the perturbation corollary extends it only through an explicit TV neighborhood. We do not establish sharp dependence on vanishing error, a uniform near-critical window, value-adaptive coordinate ordering, or memory bounds for trained neural generators. The tests do not demonstrate a hardware speedup. Within the stated interface, however, the conclusions are complete: exact generation admits a simultaneous cutwise optimum, critical approximation costs one quarter of a logarithm in bits, and exact complexity can be discontinuous under statistically negligible changes.

AI USE STATEMENT

Generative AI was used throughout this work to formulate and refine the research question and mathematical model, propose claims, develop and write proofs, search and organize related literature, design numerical diagnostics, implement and test code, interpret results, create plots, and draft and edit the manuscript. The reported numerical results were produced by execution of the supplied code, not by invented measurements. Verification in this work consists of the written derivations, finite enumeration checks, exact rational determinant calculations, and numerical consistency checks described in the paper. It is not a formal proof certificate, an independent human review, or evidence of author verification that has not occurred. Human authors are responsible for verifying and approving all claims, citations, code, and disclosures before submitting or publishing this manuscript.

REPRODUCIBILITY STATEMENT

Appendices A–D contain complete proofs. Appendix F specifies the finite experiments. The supplementary package includes the generator implementation, 52 unit tests, the seeded experiment driver, all numerical tables, plotting code, and build instructions. Mathematical statements do not depend on an inaccessible computational oracle or on a numerical PASS label.

ETHICS STATEMENT

This work studies synthetic finite probability distributions and uses no personal data or human subjects. Its principal integrity risks are overinterpreting exact rank as robust memory, treating numerical agreement as proof, and overlooking the nonuniform computation allowed by the interface. The manuscript explicitly separates these notions and discloses the scope of AI assistance.

REFERENCES

Sourav Chatterjee and Qi-Man Shao. Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *The Annals of Applied Probability*, 21(2):464–483, 2011. doi: 10.1214/10-AAP712.

Or Cohen, Vladimir Rittenberg, and Tridib Sadhu. Shared information in classical mean-field models. *Journal of Physics A: Mathematical and Theoretical*, 48(5):055002, 2015. doi: 10.1088/1751-8113/48/5/055002.

J. Hubbard. Calculation of partition functions. *Physical Review Letters*, 3:77–78, 1959. doi: 10.1103/PhysRevLett.3.77.

Rahul Jain, Yaoyun Shi, Zhaohui Wei, and Shengyu Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. In *Proceedings of the Twenty-Fourth Annual ACM–SIAM Symposium on Discrete Algorithms*, pages 1503–1512, 2013. URL `https://www.comp.nus.edu.sg/~rahul/allfiles/QStateGen.pdf`.

Frederic Koehler, Holden Lee, and Andrej Risteski. Sampling approximately low-rank Ising models: MCMC meets variational methods. In *Proceedings of the Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 4945–4988, 2022. URL `https://proceedings.mlr.press/v178/koehler22a.html`.

I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. doi: 10.1137/090752286.

Anej Svete and Ryan Cotterell. Recurrent neural language models as probabilistic finite-state automata. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8069–8086. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.emnlp-main.502.

Johannes Wilms, Julien Vidal, Frank Verstraete, and Sébastien Dusuel. Finite-temperature mutual information in a simple phase transition. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(01):P01023, 2012. doi: 10.1088/1742-5468/2012/01/P01023.

A EXACT CUTWISE REALIZATION

This appendix proves Theorem 2, including positivity, normalization, and the fact that one generator attains all cutwise bounds. The proof is valid for negative coupling, negative or zero weights, repeated sums, and arbitrary finite external fields.

A.1 A RANK LOWER BOUND FOR RETAINED STATES

Lemma 7 (Cut factorization)**.** *For any generator in Definition 1, the matrix $Q_t(u, v) = Q_T(X_{\le t} = u, X_{>t} = v)$ has ordinary rank at most $|\operatorname{supp}(S_t)|$.*

Proof. Fix a state s of positive probability. Every subsequent kernel depends on the past only through s and the known clock. All later randomness is fresh. Thus the conditional law of the remaining outputs given $(X_{\le t} = u, S_t = s)$ depends only on s. Summing over the states gives (7), a sum of $|\operatorname{supp}(S_t)|$ nonnegative rank-one matrices. States of probability zero can be discarded. This proves the rank bound. It also proves that applying arbitrary deterministic functions separately to the past and the future yields a mixture of no more than $|\operatorname{supp}(S_t)|$ product laws. □

Lemma 8 (Exponential evaluation rank)**.** *Let $u_1, \dots, u_p$ be distinct real numbers, let $v_1, \dots, v_q$ be distinct real numbers, and let $\gamma \neq 0$. Then the matrix $(e^{\gamma u_i v_j})_{i,j}$ has rank $\min(p, q)$.*

Proof. We first prove that a nonzero sum of r real exponentials with distinct exponents and nonzero coefficients has at most $r - 1$ distinct real zeros. The assertion is immediate for $r = 1$. Order the exponents as $\lambda_1 < \cdots < \lambda_r$ and divide the sum by the everywhere-positive function $e^{\lambda_1 z}$. If the result has r distinct zeros, its derivative has at least $r - 1$ distinct zeros by Rolle's theorem. That derivative is a sum of $r - 1$ exponentials with distinct exponents and nonzero coefficients, contradicting the induction hypothesis. This argument also excludes identically zero nontrivial sums. Sums with some zero coefficients are handled by using their actual number of nonzero terms.

Take any square submatrix of order $r = \min(p, q)$. A linear dependence among its columns would give a nontrivial combination of at most r exponentials $e^{\gamma v_j z}$ vanishing at r distinct points u_i. Since $\gamma \neq 0$, their exponents are distinct, contradicting the zero bound. The square submatrix is nonsingular, proving the claim. □

For a prefix configuration u and suffix configuration v, write $a(u) = \sum_{i \le t} w_i u_i$ and $c(v) = \sum_{i > t} w_i v_i$. The Gibbs matrix satisfies

$$P_t(u, v) = Z^{-1} R_t(u)\, e^{\beta a(u) c(v)/n}\, C_t(v), \tag{21}$$

$$R_t(u) = \exp\left\{ \frac{\beta a(u)^2}{2n} + \sum_{i \le t} h_i u_i \right\}, \qquad C_t(v) = \exp\left\{ \frac{\beta c(v)^2}{2n} + \sum_{i > t} h_i v_i \right\}. \tag{22}$$

All diagonal factors are strictly positive. Repeated sums yield duplicate rows or columns after these factors are removed. Selecting one representative of each distinct sum gives the compressed exponential matrix in (8). Therefore the full matrix and the compressed matrix have the same rank, namely $\min(a_t, c_t)$ by Lemma 8. Lemma 7 proves the exact lower bound at every cut.

A.2 MESSAGES AND NORMALIZED TRANSITIONS

Let G_t, F_t be the positive messages in (9) on their attainable domains; put $G_t(c) = 0$ off $\mathcal{C}_t$. Then $G_n(0) = 1$ and

$$G_t(c) = \sum_{x \in \{-1, +1\}} e^{h_{t+1} x} G_{t+1}(c - w_{t+1} x). \tag{23}$$

Splitting the suffix by its first spin likewise gives

$$F_t(a) = \sum_{x \in \{-1, +1\}} e^{h_{t+1} x} F_{t+1}(a + w_{t+1} x). \tag{24}$$

These are finite sums; no computability or time bound is being inferred from them.

Adding one weight changes a prefix support to the union of two translates of the previous support. Hence $a_{t+1} \geq a_t$. Removing one weight from a suffix has the reverse monotonicity, so $c_{t+1} \leq c_t$. Thus $\{t : a_t \leq c_t\}$ is an initial interval containing zero. Let t_* be its largest element. If $t_* = n$, then $a_n = 1$ and every weight is zero: two configurations differing in a single coordinate would otherwise have different total sums. In that case the law is a product and one state suffices at every time, consistent with the theorem. In what follows assume $t_* < n$.

For $t \leq t_*$, the state is a value of the prefix sum $a \in \mathcal{A}_t$. For $t \geq t_* + 1$, it is a suffix total $c \in \mathcal{C}_t$. Starting from $a = 0$, use (10) for $t < t_*$, (11) for $t = t_*$, and (12) for $t > t_*$. Equation (24) normalizes prefix transitions. Summing (11) over (x, c) gives exactly $F_t(a)/F_t(a) = 1$. Equation (23) normalizes suffix transitions. Every positive-probability state has a positive denominator. The last suffix state is necessarily 0.

A.3 Correctness of the full output law

Given a realized prefix u with sum a, the target conditional probability of a suffix v is

$$P(v \mid u) = \frac{\exp\{\sum_{i>t} h_i v_i + \beta(a + \sum_{i>t} w_i v_i)^2/(2n)\}}{F_t(a)}. \tag{25}$$

In particular, the future depends on the prefix only through a. Formula (10) is its singleton marginal, so the generated prefix through layer t_* has exactly the target law.

At the switch, (11) is the joint marginal of the next spin and the remaining suffix total in (25). Conditional on the sampled total c after that transition, any remaining suffix v has probability

$$\frac{e^{\sum_{i>t_*+1} h_i v_i}}{G_{t_*+1}(c)} \mathbf{1}\left\{\sum_{i>t_*+1} w_i v_i = c\right\}. \tag{26}$$

The quadratic factor and all dependence on the old prefix have canceled. Successive transitions (12) produce exactly this constrained field-weighted law: their numerators and denominators telescope, terminating with $G_n(0) = 1$. Combining the exact prefix, switch marginal, and constrained suffix proves that the full output law is $P_{\beta,w,h}$.

Because the target has full support, every value of $\mathcal{A}_t$ occurs with positive probability during the prefix phase. At and after the switch every value of $\mathcal{C}_t$ also occurs with positive probability: each has at least one suffix realization of positive target probability, and the switch samples its correct marginal. The support sizes are therefore exactly a_t for $t \leq t_*$ and c_t for $t \geq t_* + 1$. By the definition of t_* these equal $\min(a_t, c_t)$ at every cut. This completes Theorem 2 for $\beta \neq 0$. For $\beta = 0$, independent field-biased spins give width one directly.

A.4 Interpretation and a universal maximal width

The middle-cut upper bound $2^{\lfloor n/2 \rfloor}$ is not specific to Gibbs distributions. For any law on n bits, retain the full prefix up to $t = \lfloor n/2 \rfloor$. At the next transition jointly sample the next bit and the entire remaining suffix conditional on that prefix, then successively emit and delete the retained suffix. The largest retained alphabet has size at most $2^{\lfloor n/2 \rfloor}$. Thus the injective-weight examples attain the maximal width permitted by this interface among all n-bit laws.

The representation switch is important even for nonuniform computation: it avoids retaining both a prefix sum and a sampled suffix sum between emissions. They may both be involved in one abstract transition, whose temporary arithmetic is not charged by Definition 1. Charging such temporary computation, transition-table size, or the clock would define a different resource problem.

B Critical lower bound at fixed total variation

We prove the finite-n estimates used in Section 4. Constants denoted by c, C may change between displays, but are positive and independent of n. All statements in this appendix use even $n \geq 2$ and zero external field.

B.1 A UNIFORM CRITICAL POINT-MASS BOUND

Let $K = (M+n)/2$ be the total number of $+1$ spins. At $\beta = 1$ its unnormalized mass at magnetization $M \in \{-n, -n+2, \dots, n\}$ is

$$W_n(M) = \binom{n}{(n+M)/2} e^{M^2/(2n)}. \tag{27}$$

Define, with the endpoint values taken by continuity,

$$I(s) = \frac{1+s}{2}\log(1+s) + \frac{1-s}{2}\log(1-s), \qquad |s| \le 1. \tag{28}$$

Integrating $I'(s) = \operatorname{arctanh}(s)$ gives the positive-coefficient series

$$I(s) = \sum_{j=1}^{\infty} \frac{s^{2j}}{(2j-1)(2j)} \ge \frac{s^2}{2} + \frac{s^4}{12}. \tag{29}$$

For $|s| \le 1/2$, the same series gives $I(s) \le s^2/2 + Cs^4$.

Stirling's inequalities imply that, uniformly for $|M| \le n/2$,

$$c\frac{2^n}{\sqrt{n}} e^{-nI(M/n)} \le \binom{n}{(n+M)/2} \le C\frac{2^n}{\sqrt{n}} e^{-nI(M/n)}. \tag{30}$$

For completeness, insert $c_0\sqrt{k}(k/e)^k \le k! \le C_0\sqrt{k}(k/e)^k$ into $n!/[k!(n-k)!]$ for $k, n-k \ge n/4$. The square-root prefactor is between constant multiples of $n^{-1/2}$; the remaining exponential is $\exp\{n\log 2 - nI(M/n)\}$. The finitely many small cases are absorbed in the constants.

In the central region, (29) and (30) give $W_n(M) \le C2^n/\sqrt{n}$. For the outer region, use $\binom{n}{k} \le \exp\{nH(k/n)\}$, valid also at the endpoints, where $H((1+s)/2) = \log 2 - I(s)$. One proof is to take the kth term in the binomial expansion of $1 = (p+(1-p))^n$ with $p = k/n$. Since $|M/n| > 1/2$,

$$W_n(M) \le 2^n e^{-n(M/n)^4/12} \le 2^n e^{-n/192} \le C\frac{2^n}{\sqrt{n}}. \tag{31}$$

Thus the last bound holds for every admissible magnetization.

For all sufficiently large n, the interval $|M| \le n^{3/4}/4$ lies in $|M| \le n/2$. It contains at least $cn^{3/4}$ admissible values with the correct parity. By the lower Stirling bound and $I(s) \le s^2/2 + Cs^4$, each corresponding $W_n(M)$ is at least $c2^n/\sqrt{n}$. Consequently

$$\sum_M W_n(M) \ge c2^n n^{1/4}. \tag{32}$$

Increasing constants handles the remaining finite values of n. We have proved

$$\max_{0\le k\le n} P_{n,1}(K=k) \le Cn^{-3/4}. \tag{33}$$

This is a finite-size anti-concentration bound. It does not assume a limiting density theorem.

B.2 MARGINALIZING TO A HALF-SYSTEM

Write $n = 2m$, $A = \sum_{i\le m} X_i$, $B = \sum_{i>m} X_i$, and $J = (A+m)/2$. Conditional on $K = k$, J is hypergeometric, with

$$H_{j,k} := P(J = j \mid K = k) = \frac{\binom{m}{j}\binom{m}{k-j}}{\binom{n}{k}}, \tag{34}$$

where binomial coefficients outside their natural range are zero. Its row sums obey

$$\sum_{k=0}^{n} H_{j,k} = \frac{n+1}{m+1} < 2. \tag{35}$$

To prove the identity, sample p uniformly on $[0,1]$ and, conditional on p, draw n independent Bernoulli variables. The beta integral gives $P(K=k) = \binom{n}{k}\int_0^1 p^k(1-p)^{n-k}dp = 1/(n+1)$. Similarly, $P(J=j) = 1/(m+1)$. Conditioning the uniform-p mixture on its total count makes all strings with that count equiprobable, yielding the same hypergeometric kernel. Multiplying its row sum by $1/(n+1)$ gives the identity.

Equations (33) and (35) imply

$$P_{n,1}(J=j) = \sum_k H_{j,k}P_{n,1}(K=k) \le 2Cn^{-3/4}. \tag{36}$$

The spin sum A is on a lattice of spacing two. Any half-open interval of length $W\sqrt{n}$ therefore contains at most $W\sqrt{n}/2+1$ possible values. Hence

$$\sup_{a\in\mathbb{R}} P_{n,1}(A \in [a, a+W\sqrt{n})) \le C(W+1)n^{-1/4}, \qquad W>0. \tag{37}$$

B.3 The discrepancy between the halves

Conditional on M, exchangeability gives $\mathbb{E}[X_i \mid M] = M/n$. For $i \neq j$,

$$\mathbb{E}[X_iX_j \mid M] = \frac{M^2-n}{n(n-1)}, \qquad \mathrm{Cov}(X_i, X_j \mid M) = -\frac{1-(M/n)^2}{n-1}. \tag{38}$$

The second formula follows by subtracting $(M/n)^2$ from the first. Therefore

$$\mathrm{Var}(A \mid M) = \left(1-\frac{M^2}{n^2}\right)\left(m - \frac{m(m-1)}{n-1}\right) = \frac{n^2-M^2}{4(n-1)}. \tag{39}$$

Since $\mathbb{E}[A \mid M] = M/2$ and $A-B = 2A-M$,

$$\mathbb{E}[(A-B)^2 \mid M] = \frac{n^2-M^2}{n-1} \le 2n, \qquad \mathbb{E}|A-B| \le \sqrt{2n}. \tag{40}$$

This estimate uses exchangeability, not criticality.

B.4 Quantization and a support-cardinality converse

Fix an interval length $\ell = W\sqrt{n}$ and an offset $\theta \in [0,\ell)$. Let

$$U_\theta = \lfloor (A+\theta)/\ell \rfloor, \qquad V_\theta = \lfloor (B+\theta)/\ell \rfloor. \tag{41}$$

Although the floor function ranges over integers, only finitely many values occur here. For any two real numbers a, b, a uniform offset separates their bins with probability $\min(1, |a-b|/\ell)$. Indeed, if their distance is smaller than ℓ, separation occurs exactly when a grid boundary lies between them; the set of offsets has length $|a-b|$. If the distance is at least ℓ, separation occurs for every offset, up to zero-measure boundary cases.

Averaging under the target and using (40),

$$\frac{1}{\ell}\int_0^\ell P(U_\theta \neq V_\theta)\, d\theta \le \frac{\mathbb{E}|A-B|}{\ell} \le \frac{\sqrt{2}}{W}. \tag{42}$$

There is consequently a fixed offset with $\delta := P(U_\theta \neq V_\theta) \le \sqrt{2}/W$. The marginal maximum for every offset is at most $\alpha = C(W+1)n^{-1/4}$ by (37).

Now let a generator produce an arbitrary law Q_T with $\mathrm{TV}(P_{n,1}, Q_T) \le \varepsilon$. Under this law, the quantized half-sums form a mixture of at most $|\operatorname{supp}(S_m)|$ product laws by Lemma 7. Their distance from the target quantized law is at most ε: for any event in quantized space, its preimage is an event in the original space, which proves TV contraction. Lemma 4, proved in the main text, yields

$$|\operatorname{supp}(S_m)| \ge \frac{1-2\sqrt{2}/W - 2\varepsilon}{C(W+1)}n^{1/4}. \tag{43}$$

Set $W=16$ and use $\varepsilon \le 1/8$. The numerator is at least $3/4 - \sqrt{2}/8 > 0$. Absorbing fixed factors into c proves the lower bound in (5). This argument allows arbitrary hidden-state randomization and arbitrary nonsymmetric approximating laws. It does not use continuity of mutual information on a growing alphabet.

C AUXILIARY-FIELD UPPER BOUNDS IN ALL THREE REGIMES

We prove the upper bounds in Theorem 3 in the exact retained-state interface. Numerical approximations to the kernels are not assumed in this proof.

C.1 A MIXTURE IDENTITY AND ITS METRIC

For $\beta > 0$, the Gaussian moment-generating identity gives

$$e^{\beta M^2/(2n)} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-z^2/2+\sqrt{\beta/n}zM}\, dz. \tag{44}$$

Summing over spin vectors normalizes this identity and yields (15). The latent density is integrable because $\log\cosh t \le |t|$, so its log density is bounded above by $-z^2/2 + \sqrt{n\beta}|z|$.

Let $\psi(z) = n\log[2\cosh(\sqrt{\beta/n}z)]$ be the log-partition function of the conditional product family. Then

$$\psi''(z) = \beta\,\mathrm{sech}^2(\sqrt{\beta/n}z) \le \beta. \tag{45}$$

The exponential-family divergence formula and Taylor's theorem imply

$$D_{\mathrm{KL}}(P_z\|P_{z'}) = \psi(z') - \psi(z) - \psi'(z)(z'-z) \le \frac{\beta}{2}(z'-z)^2. \tag{46}$$

Pinsker's inequality, $\mathrm{TV}(P,Q) \le \sqrt{D_{\mathrm{KL}}(P\|Q)/2}$, gives (16). Here Pinsker's inequality can also be obtained by mapping to an event achieving TV, applying the log-sum inequality to the event and its complement, and using the scalar binary divergence bound $d(p\|q) \ge 2(p-q)^2$.

Lemma 9 (Latent covering)**.** *Suppose a union of L bounded intervals of total length Λ has $\pi_{n,\beta}$-mass at least $1-\varepsilon/2$. There is a generator of width at most*

$$L + \frac{\sqrt{\beta}\,\Lambda}{\varepsilon} \tag{47}$$

rounded up to an integer, whose output law is within ε of $P_{n,\beta}$.

Proof. Partition each interval into at most $\lceil \text{length}\sqrt{\beta}/\varepsilon \rceil$ cells and choose a representative in each cell. First condition the latent law on the union, an operation changing it in TV by exactly its outside mass. Next map the conditioned latent to the cell representative. Every displacement is at most $\varepsilon/\sqrt{\beta}$. By the mixture identity, TV contraction for the first operation and convexity/coupling for the second, the output law changes by at most $\varepsilon/2 + \varepsilon/2 = \varepsilon$. Explicitly, under a coupling of a latent and its representative, the conditional output-TV bound is (16); averaging bounds the mixture TV.

Let the finitely many representatives be $z_1, \dots, z_K$ with weights $\lambda_1, \dots, \lambda_K$. At the first transition jointly draw (X_1, S_1) with probability $\lambda_j P_{z_j}(X_1 = x)$. At later nonterminal transitions preserve j and draw the next spin according to P_{z_j}. On the final transition drop the index. The output is exactly $\sum_j \lambda_j P_{z_j}$, and at most K states are retained. The cell count is at most the ceiling of (47). This construction includes the cost of the persistent mixture label; no shared seed is free. □

C.2 SUBCRITICAL CONCENTRATION

Let

$$V_{n,\beta}(z) = \frac{z^2}{2} - n\log\cosh(\sqrt{\beta/n}z), \qquad \pi_{n,\beta}(dz) = \frac{e^{-V_{n,\beta}(z)}dz}{\int_{\mathbb{R}} e^{-V_{n,\beta}(u)}du}. \tag{48}$$

The elementary bound $0 \le \log\cosh t \le t^2/2$ follows by integrating $0 \le \tanh t \le t$ for $t \ge 0$ and using evenness. Thus for $0 < \beta < 1$,

$$\frac{1-\beta}{2}z^2 \le V_{n,\beta}(z) \le \frac{z^2}{2}. \tag{49}$$

The denominator in (48) is at least $\sqrt{2\pi}$, and

$$\mathbb{E}_\pi z^2 \le \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^2 e^{-(1-\beta)z^2/2}dz = (1-\beta)^{-3/2}. \tag{50}$$

Taking a centered interval of radius $(2/\varepsilon)^{1/2}(1-\beta)^{-3/4}$ leaves mass at most $\varepsilon/2$ outside. Lemma 9 gives width at most $1 + C_\beta \varepsilon^{-3/2}$, absorbed into (6) for $\varepsilon \le 1/8$.

C.3 SUPERCRITICAL CONCENTRATION AROUND MOVING WELLS

Fix $\beta > 1$ and write $z = \sqrt{n}\, t$. Then $V_{n,\beta}(z) = n v_\beta(t)$ with

$$v_\beta(t) = t^2/2 - \log\cosh(\sqrt{\beta} t), \qquad v'_\beta(t) = t - \sqrt{\beta}\tanh(\sqrt{\beta} t). \tag{51}$$

For $u > 0$, $\tanh u/u$ decreases strictly from 1 to 0. Indeed its derivative has the sign of $u\operatorname{sech}^2 u - \tanh u$, which is negative because $g(u) = \tanh u - u\operatorname{sech}^2 u$ has $g(0) = 0$ and derivative $2u\operatorname{sech}^2 u \tanh u > 0$.

The equation $v'_\beta(t) = 0$ therefore has precisely one positive solution t_β and its negative, besides zero. The origin is a strict local maximum, since $v''_\beta(0) = 1 - \beta < 0$. The other two points are the only minima and are global, since $v_\beta(t) \to \infty$ as $|t| \to \infty$. At $u = \sqrt{\beta} t_\beta$, the defining equation is $\beta = u/\tanh u$, so

$$v''_\beta(t_\beta) = 1 - \beta\operatorname{sech}^2 u = 1 - \frac{u\operatorname{sech}^2 u}{\tanh u} > 0. \tag{52}$$

Let $d_\beta(t) = \operatorname{dist}(t, \{-t_\beta, t_\beta\})$ and $v_* = v_\beta(t_\beta)$. There exist finite positive constants c_β, C_β such that

$$c_\beta d_\beta(t)^2 \le v_\beta(t) - v_* \le C_\beta d_\beta(t)^2 \quad (t \in \mathbb{R}). \tag{53}$$

For the lower bound, the ratio on the left extends continuously and positively at either minimum, with limit $v''_\beta(t_\beta)/2$; away from those points its numerator is strictly positive. At infinity its limit is $1/2$, since $\log\cosh(\sqrt{\beta} t) = O(|t|)$. Compactness then gives a positive infimum. The upper bound follows by the same argument or by the bounded second derivative and Taylor's theorem around the nearest minimum.

Put $D_n(z) = \operatorname{dist}(z, \{\pm\sqrt{n}\, t_\beta\})$. Equation (53) gives

$$c_\beta D_n(z)^2 \le V_{n,\beta}(z) - n v_* \le C_\beta D_n(z)^2. \tag{54}$$

The normalizing integral after subtracting $n v_*$ is bounded below by $2e^{-C_\beta}$, by integrating over a unit-radius interval about either center. Its second-distance numerator is bounded above by

$$\int_{\mathbb{R}} D_n(z)^2 e^{-c_\beta D_n(z)^2} dz \le 2\int_{\mathbb{R}} u^2 e^{-c_\beta u^2} du < \infty. \tag{55}$$

The inequality partitions the line into the two Voronoi regions of the centers and translates each region to its center. We obtain the dimension-uniform estimate

$$\mathbb{E}_\pi D_n(z)^2 \le C_\beta. \tag{56}$$

Two intervals about the centers, each of radius $(2C_\beta/\varepsilon)^{1/2}$, cover latent mass at least $1 - \varepsilon/2$. Their union has total length at most $4(2C_\beta/\varepsilon)^{1/2}$, whether or not they overlap. Lemma 9 proves (6). Notice that the increasing separation of the centers does not increase the length being covered.

C.4 CRITICAL QUARTIC CONCENTRATION

For $\beta = 1$, write $f(t) = t^2/2 - \log\cosh t$, so $V_{n,1}(z) = n f(z/\sqrt{n})$. For $t \ge 0$, the function $\tanh t - t + t^3/3$ has derivative $t^2 - \tanh^2 t \ge 0$ and vanishes at zero. Thus $\tanh t \ge t - t^3/3$ and

$$0 \le f(t) = \int_0^t (u - \tanh u) du \le t^4/12. \tag{57}$$

For the lower bound, $f(t) > 0$ for $t \ne 0$, since $\tanh t < t$ for $t > 0$. Its ratio to t^4 has limit $1/12$ at zero, by Taylor expansion of $\log\cosh t$. Its ratio to t^2 has limit $1/2$ at infinity. Positivity and compactness on intermediate intervals yield an absolute $c > 0$ such that

$$f(t) \ge c\min(t^4, t^2) \quad (t \in \mathbb{R}). \tag{58}$$

The Taylor coefficient can be checked directly from $f''(t) = \tanh^2 t$, whose ratio to t^2 tends to one; two integrations give $f(t)/t^4 \to 1/12$.

The normalizing integral is bounded below by integrating over $|z| \le n^{1/4}$, where (57) gives $V_{n,1}(z) \le 1/12$:

$$\int_{\mathbb{R}} e^{-V_{n,1}(z)} dz \ge 2e^{-1/12} n^{1/4}. \tag{59}$$

Using (58) and splitting at $|z| = \sqrt{n}$,

$$\int_{\mathbb{R}} z^2 e^{-V_{n,1}(z)} dz \le \int_{|z|\le\sqrt{n}} z^2 e^{-cz^4/n} dz + \int_{|z|>\sqrt{n}} z^2 e^{-cz^2} dz \tag{60}$$

$$\le Cn^{3/4} + C \le C' n^{3/4}. \tag{61}$$

The first integral is bounded by its full-line version and the substitution $z = n^{1/4}u$. Dividing by (59) proves

$$\mathbb{E}_\pi z^2 \le C\sqrt{n}. \tag{62}$$

A centered interval of radius $(2C/\varepsilon)^{1/2} n^{1/4}$ now has outside mass at most $\varepsilon/2$. Lemma 9 yields width at most $1 + C'\varepsilon^{-3/2} n^{1/4}$, and the additive one is absorbed in the constant. Together with Appendix B, this completes Theorem 3.

D Perturbation stability and input precision

D.1 A Gibbs perturbation bound

Lemma 10 (Uniform energy perturbation)**.** *On a finite sample space, let* $P_0(x) \propto e^{H(x)}$ *and* $P_1(x) \propto e^{H(x)+V(x)}$*. If* $\|V\|_\infty \le D$*, then* $\mathrm{TV}(P_0, P_1) \le \min(1, D)$.

Proof. Set $P_s(x) = e^{H(x)+sV(x)}/Z_s$ for $0 \le s \le 1$. Differentiating the finite sum gives $\partial_s P_s(x) = P_s(x)(V(x) - \mathbb{E}_s V)$. Therefore

$$\mathrm{TV}(P_0, P_1) \le \frac{1}{2} \int_0^1 \sum_x |\partial_s P_s(x)| ds = \frac{1}{2} \int_0^1 \mathbb{E}_s |V - \mathbb{E}_s V| ds \le D. \tag{63}$$

The universal TV bound of one completes the statement. This deliberately conservative constant is sufficient here. □

For $w_i^{(\delta)} = 1 + \delta 2^{-i}$, write $S = \sum_i x_i$ and $e = \sum_i \delta 2^{-i} x_i$. Then $|S| \le n$ and $|e| < \delta$. The Hamiltonian difference obeys

$$\left| \frac{\beta}{2n} [(S+e)^2 - S^2] \right| \le |\beta| \left(\delta + \frac{\delta^2}{2n} \right). \tag{64}$$

The fields are unchanged. Lemma 10 proves the TV part of Theorem 5 uniformly over all finite fields.

D.2 Injectivity of the perturbed subset sums

For a subset $I \subseteq [n]$, its unsigned weight sum is

$$W(I) = |I| + \delta \sum_{i \in I} 2^{-i}. \tag{65}$$

If two subsets have different cardinalities, their integer parts differ by at least one, whereas the magnitude of the perturbation difference is less than $\delta < 1$. They cannot have the same weight sum. If they have the same cardinality, equality would require equality of their finite binary sums. Multiplication by 2^n makes these sums integers whose binary expansions use distinct positions, so equality implies identical subsets. This also follows by inspecting the smallest index in the symmetric difference, whose power exceeds the sum of all subsequent powers through index n.

For any set of coordinates, the signed sum is twice the weight sum of its $+1$ subset minus the fixed total weight of those coordinates. It is therefore injective in the sign configuration. At every cut, $a_t = 2^t$ and $c_t = 2^{n-t}$, independently of a fixed coordinate permutation. Theorem 2 yields $K_0 = 2^{\lfloor n/2 \rfloor}$ and the stated exact bit count.

D.3 FIXED-ERROR TRANSFER

Suppose $\mathrm{TV}(P, P') \le r < \varepsilon$. A generator within $\varepsilon - r$ of P is within ε of P', while a generator within ε of P' is within $\varepsilon + r$ of P. Taking the minimum widths proves (18); no comparison between internal states of the two generators is required.

For the zero-field perturbation in Theorem 5, assume $\varepsilon \le 1/16$ and the displayed energy bound is at most $\varepsilon/2$. Then $\varepsilon - r \ge \varepsilon/2$ and $\varepsilon + r \le 3\varepsilon/2 \le 1/8$. For even n at $\beta = 1$, applying both sides of Theorem 3 through the sandwich gives

$$cn^{1/4} \le K_\varepsilon(P_{1,w^{(\delta)},0}) \le C(\varepsilon/2)^{-3/2} n^{1/4}. \tag{66}$$

For every fixed $\beta > 0$, $\beta \ne 1$, it gives $K_\varepsilon \le C_\beta(\varepsilon/2)^{-3/2}$. The width lower bound one completes the constant-order classification. The perturbation amplitude can be fixed small enough independently of n, since $n \ge 2$, or can tend to zero with n.

D.4 UPPER BOUND WITH A DYADIC INPUT ALPHABET

For $w_i \in 2^{-b}\mathbb{Z}$, changing the sign of one coordinate changes the signed sum by a multiple of 2^{1-b}. Thus any prefix or suffix support is contained in one coset of the lattice $2^{1-b}\mathbb{Z}$. With $|w_i| \le 2$, all such sums lie in $[-2n, 2n]$. Each support therefore has at most $2n2^b + 1$ points. It also has at most 2^t or 2^{n-t} points from counting sign configurations. Consequently

$$K_0 \le \min\{2^{\lfloor n/2 \rfloor}, 2n2^b + 1\}, \tag{67}$$

by Theorem 2. Taking the ceiling of the base-two logarithm yields (19).

D.5 A MATCHING FAMILY

Let $n = 2m$ and $r = \min(b, m)$. Form a half-list of r special weights $1 + 2^{-j}$, $j = 1, \dots, r$, and $m - r$ ordinary weights 1. Append the reversed half-list. These weights lie in $[1, 3/2]$ and on the required dyadic grid. For a subset of a half-list, the fractional part of its sum uniquely identifies the selected special weights: it is a finite binary fraction in $[0, 1)$. Once those weights are identified, the integer part identifies the number, but not the identities, of selected ordinary weights. There are exactly $(m - r + 1)2^r$ distinct unsigned sums, and the same number of signed sums.

At the middle cut both supports have this size. For $t \le m$, monotonicity gives $a_t \le a_m$, and for $t \ge m$ it gives $c_t \le c_m$. Thus the maximum of $\min(a_t, c_t)$ is attained at the middle and equals $(m - r + 1)2^r$. This proves (20), for arbitrary finite fields.

To spell out the final order statement, the universal upper bound is $O(\min\{n, b + \log_2 n\})$. If $b \ge m$, the construction needs m bits. If $m/2 \le b < m$, its logarithmic width is at least $b \ge m/2$, a constant fraction of n. If $b < m/2$, it is $b + \log_2(m - b + 1) \ge b + \log_2(m/2 + 1)$, which is within a fixed additive constant of $b + \log_2 n$. These cases, including the finitely many smallest m, prove the matching Θ lower bound uniformly for even $n \ge 2$ and integer $b \ge 0$.

E WHAT THE STATEMENTS DO AND DO NOT IMPLY

Exact arithmetic versus finite precision. The mathematical kernels are real probability tables, treated as read-only. Their entries need not be dyadic even when all weights are dyadic. Proposition 6 counts the effect of input precision on state cardinality; it does not assert a finite fair-coin implementation of each transition using the same memory. The code evaluates probabilities in floating point and is labeled accordingly.

Deterministic summaries versus stochastic states. The lower bounds allow states randomized conditional on a prefix. Such randomization can lower memory compared with storing an exact deterministic statistic. The exact upper bound explicitly uses it when it changes to a sampled suffix total. Thus a count of distinct conditional distributions alone would not be a sufficient lower-bound proof; the matrix factorization argument is required.

One cut versus the entire generator. An independent nonnegative factorization at each cut does not automatically specify compatible transitions. Theorem 2 resolves compatibility for this family through its single representation switch. Theorem 3 provides a full-sequence upper bound by retaining the same discrete latent index across all interior layers. Its lower bound needs only the balanced cut.

Temperature and error limits. For exact sampling, the limit $\beta \to 0$ is discontinuous at zero whenever the weights produce dependence for $\beta \neq 0$. For approximate sampling, our constants may depend on a fixed β and fixed error. Taking $\beta = \beta_n \to 1$ or $\varepsilon = \varepsilon_n \to 0$ is not licensed by a fixed-parameter asymptotic statement. The explicit $\varepsilon^{-3/2}$ upper bounds remain valid within their stated quantifiers, but do not provide matching small-error lower bounds.

Scope of novelty. The exponential zero-count argument, Gaussian transform, nonnegative-rank interface, and critical fluctuation mechanism have classical antecedents. The paper's contribution is their operational synthesis into a simultaneous exact-width realization, a fixed-TV support-cardinality classification, and explicit exact-versus-approximate perturbation and precision consequences. In particular, neither the existence of critical shared information nor its $1/4$ logarithmic coefficient is claimed as a new physical phenomenon.

F Complete experimental specification

The supplementary archive contains `code/core.py`, `code/experiments.py`, `code/plots.py`, `tests/test_core.py`, and the raw CSV and JSON results. All experiments are deterministic given the stated seed and software arithmetic. There are no training sets, learned parameters, GPUs, or omitted training runs.

F.1 Exact state-machine checks

The 150 enumerated cases comprise all combinations of $n = 1, \dots, 10$, $\beta \in \{-1.5, 0, 0.5, 1, 2\}$, and the following weight choices. Homogeneous weights are all one. Dyadic weights are $w_i = 1 + 2^{-(i+3)}$, with i starting at one. Signed weights are drawn from $\{-4, -3, \dots, 4\}/4$. Fields are drawn uniformly from $[-0.6, 0.6]$. One NumPy random generator, seeded with 20260915, supplies the signed weights and fields in the loop order given in `experiments.py`; no best seed is selected.

The direct enumerator evaluates the normalized energy in (1) for all 2^n vectors. A separate routine constructs the suffix messages in (23) and prefix messages in (9), then propagates the probabilities of pairs (emitted prefix, retained state) through the three kinds of transition. It sums over hidden states only at the end. Sum labels use Python `Fraction`; logarithms of message masses and probabilities use binary64 with log-sum-exp stabilization. At $\beta = 0$ it uses the explicit independent-spin kernel.

The comparison records full-law TV, maximum absolute atom error, terminal normalization residual, local transition-normalization residual, the peak state count, and equality of the entire width profile to the theorem. All cases pass the declared 2×10^{-11} numerical tolerances. The observed maximum TV is $1.0318436778839271 \times 10^{-15}$ and the largest local residual is $2.886579864025407 \times 10^{-15}$.

As an additional exact finite arithmetic check, the program forms the 4×4 matrix

$$H_q(a, c) = q^{(a+c)^2}, \qquad a, c \in \{-3, -1, 1, 3\}, \quad q \in \{3/2, 2/3\}. \tag{68}$$

Gaussian elimination with rational arithmetic gives a nonzero determinant for each matrix. The full fractions are recorded in `rational_determinants.json`. This supplies two finite rank witnesses, one of each coupling sign, not a proof about arbitrary dimension.

F.2 Exchangeable count-law evaluation

For the homogeneous target the count distribution is evaluated directly as

$$p_k = \frac{\binom{n}{k} \exp\{\beta(2k - n)^2/(2n)\}}{\sum_{j=0}^{n} \binom{n}{j} \exp\{\beta(2j - n)^2/(2n)\}}, \qquad k = 0, \dots, n. \tag{69}$$

A discrete latent mixture yields

$$q_k = \sum_{a=1}^{K} \lambda_a \binom{n}{k} p_a^k (1-p_a)^{n-k}, \qquad p_a = \frac{1+\tanh(\sqrt{\beta/n}\, z_a)}{2}. \tag{70}$$

Both vector distributions are uniform within a count class. For a vector with k positive spins their probabilities are $p_k/\binom{n}{k}$ and $q_k/\binom{n}{k}$. Summing over the $\binom{n}{k}$ vectors in each class proves the exact identity

$$\mathrm{TV}(P,Q) = \frac{1}{2}\sum_{k=0}^{n} |p_k - q_k|. \tag{71}$$

Thus these experiments measure the full-vector error, not merely a weaker marginal discrepancy. The computation uses log-factorials via `gammaln`, log-sum-exp, and chunked binomial-mixture evaluation. A final floating-point normalization of q is recorded rather than hidden.

F.3 Latent integration and mesh selection

The grid uses $\beta \in \{0.5, 1, 1.5\}$ and $n \in \{64, 128, 256, 512, 1024, 2048, 4096, 8192\}$. Integration regions are prescribed as follows. Below criticality, the region is $[-8/\sqrt{1-\beta}, 8/\sqrt{1-\beta}]$. At criticality, it is $[-6n^{1/4}, 6n^{1/4}]$. Above criticality, compute the positive solution $m_\beta = \tanh(\beta m_\beta)$ with Brent's method, set $c_n = \sqrt{n\beta}\, m_\beta$, and use radius $R = 10/\sqrt{1-\beta(1-m_\beta^2)}$ about each center $\pm c_n$. Overlapping intervals are merged. These are numerical integration choices; the program does not claim interval-certified tail masses.

For requested mesh s, split each interval into $\lceil \text{length}/s \rceil$ equal cells. The representative is the cell midpoint. Integrate the stabilized log latent density inside each cell with Gauss–Legendre quadrature and normalize the cell weights over the declared region. The fixed-mesh experiment uses $s = 0.25$ and orders 8 and 16 for a consistency comparison. The largest computed TV from the target is 0.00031153766998437757, and the largest TV between the order-8 and order-16 approximations is $1.7116765660589546 \times 10^{-14}$.

The fixed-error experiment uses all eight meshes $\{0.25, 0.5, 1, 2, 4, 8, 16, 32\}$ with quadrature order 32, for a total of 192 candidate approximations. For each (β, n) it selects the smallest state count whose computed target TV is at most 0.05, breaking ties by TV. All candidates, not just selected ones, are retained in `fixed_error_candidates.csv`. This finite search does not imply global optimality over meshes, interval choices, mixture weights, or general generators. Selected mesh 2 is feasible and has the smallest width among the tested candidates in every reported case.

Table 2: Selected widths and actual count-law TVs at three representative sizes. The candidate family and tolerance are fixed as above.

β	n	States	TV	Requested mesh
0.5	64	12	0.0174144	2
0.5	1024	12	0.0174501	2
0.5	8192	12	0.0174518	2
1	64	17	0.0080068	2
1	1024	34	0.0053853	2
1	8192	58	0.0041238	2
1.5	64	22	0.0171680	2
1.5	1024	26	0.0179355	2
1.5	8192	26	0.0179425	2

F.4 Critical lower-witness diagnostics

For even $n \in \{64, 128, 256, 512, 1024, 2048\}$, the exact-form balanced count matrix is

$$P(J=j, L=\ell) \propto \binom{m}{j}\binom{m}{\ell} \exp\left\{\frac{[2(j+\ell)-n]^2}{2n}\right\}, \qquad m = n/2. \tag{72}$$

We evaluate it numerically and quantize both spin sums using a common interval width $W\sqrt{n}$. The diagnostic searches $W \in \{1, 2, 3, 4, 6, 8, 12, 16\}$ and offsets at the 12 equally spaced fractions of one bin. At $\varepsilon = 0.05$, it records α, δ, and the expression $\max\{0, (1 - 2\delta - 2\varepsilon)/\alpha\}$ from Lemma 4, selecting the largest expression within this reported family.

The six selected expressions are approximately 1.103, 1.216, 1.311, 1.392, 1.459, and 1.514. They are weak finite bounds and do not numerically resolve the asymptotic exponent on this range. The asymptotic conclusion comes from the uniform anti-concentration proof and offset-averaging argument, not from fitting these values. Figure 3 separately shows the critical point-mass scaling directly involved in that proof.

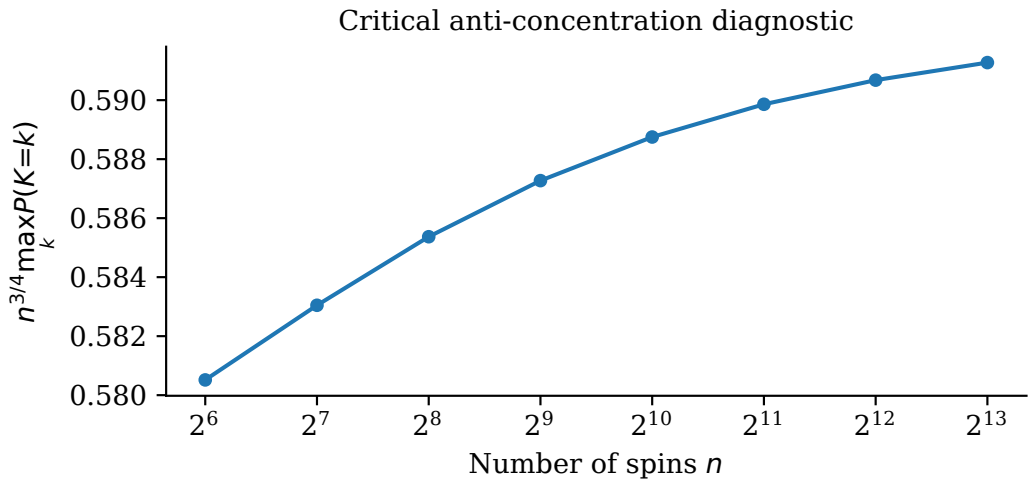

Figure 3: A finite diagnostic of the anti-concentration ingredient: $n^{3/4}$ times the largest target count probability. This is an exact-form finite sum evaluated in floating point, not an asymptotic proof or interval bound.

F.5 Perturbation diagnostics and software environment

The perturbation grid uses $n = 10$, $\beta = 1.5$, zero fields, and $\delta = 2^{-q}$ for $q \in \{2, 4, 6, 8, 10, 12, 14, 16\}$. All state support calculations are exact rational set operations. Both vector distributions are fully enumerated to compute TV. The smallest-amplitude TV is $3.8113579593910185 \times 10^{-6}$; every distance lies below the analytical bound in Theorem 5.

The recorded environment is Python 3.13.5, NumPy 2.3.5, SciPy 1.17.0, and Matplotlib 3.10.8. The test suite has 52 cases. The experiment driver records its local elapsed time and exact software versions in `results/summary.json`; elapsed time is an execution record, not a performance comparison. The files `audit/tests_green.txt` and `audit/experiment_run.txt` preserve the run output. No result depends on access to a private dataset, a remote API, or a trained model.