

Gibbs Approximation from Coarse-Graining of Finite Systems

Sun Zhenyu^{1*} and Chen Chen²

^{1*}School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao, 066004, China.

²Sanli Honors College, School of Physics and Electronics Engineering, Shanxi University, Taiyuan, Shanxi, 030006, China.

*Corresponding author: sunzhenyu@neuq.edu.cn

Abstract

Thermodynamic integrability in finite systems admits concave probability-response potentials beyond the canonical ensemble. We investigate how replacing an unresolved block of energy levels by its effective free energy constrains this class and determines the Gibbs distribution, both exactly and quantitatively. For a response family defined across all finite spectra, exact single-block replacement implies conditional independence within each block and selects the Gibbs family with its canonical free-energy normalization. On a bounded energy window, conditional log-odds discrepancies determine a common reference inverse-temperature scale and yield a dimension-independent total-variation bound. A discrete calibration protocol incorporates grid rounding, interpolation under stated regularity bounds, and finite-sample concentration guarantees. We bound the conditional log-odds discrepancies in terms of coarse-grained block-probability residuals through finite block suppression, a bounded outer-control displacement, and a local internal curvature bound. The estimate separates auxiliary system size, measurement precision, and probability conditioning. Smooth local perturbations show that finitely many observations do not uniquely determine the global response. Applications to finite-capacity reservoirs and fixed-coupling mean-force models distinguish work-form closure from temperature-energy scaling. For two subsystems sharing small, bounded, symmetric log-temperature fluctuations, the leading mutual information is half the squared log-temperature variance times the product of their reference canonical heat capacities.

Keywords: Gibbs distribution, coarse-graining, thermodynamic integrability, probability ratios, finite reservoirs, temperature fluctuations

1 Introduction

The canonical Gibbs distribution describes the equilibrium populations of a system weakly coupled to a thermal reservoir. Thermodynamic integrability constrains their response to changes in the energy levels by requiring the existence of free energy and entropy. For finite systems, these integrability conditions also admit noncanonical probability responses. We ask whether consistency under coarse-graining selects the Gibbs distribution, and what quantitative conclusion remains when system sizes, energy ranges, and measurement precision are finite.

Zhai and Sun [1] formulated the entropy-integrability equations (EIE) for finite diagonal systems and constructed canonical and noncanonical solutions, including temperature-scale mixtures and

correlations induced by a shared fluctuating temperature. We distinguish the closure condition for the reversible work differential from the scaling relation between temperature and energy. Finite-dimensional convex duality [2, 3] associates each spectrally stable response potential with a concave thermodynamic entropy, without prescribing the Shannon form.

Several complementary approaches in statistical physics, information theory, and decision theory characterize exponential weights through consistency or independence principles. Classical axiomatizations of Shannon entropy rely on recursive grouping identities [4–6], which connect to equilibrium statistical mechanics via Jaynes’s maximum-entropy variational principle [7]. Directly at the level of state probabilities, Cerreia-Vioglio et al. [8] introduced axiomatic tests for the Boltzmann distribution. In the language of stochastic choice, where utility corresponds to negative energy, Sandomirskiy and Tamuz [9] characterized the multinomial logit rule and its quantitative stability under continuity, decomposability, and neutrality, and established a related characterization of the Boltzmann law via convolutional independence of energy distributions [10]. A later revision by Sandomirskiy, Sung, Tamuz, and Wincelberg, *A Separability Foundation for Random Coefficients Logit* (arXiv:2312.04827v4, 2026), characterized random-coefficients mixtures of linear logit models under marginal consistency, monotonicity, and continuity. Fu et al. [11] derived exponential weights from invariance under coarse-graining scale transformations and uniform energy shifts.

The coarse-graining operation studied in the present work replaces a selected block of energy levels by the value of its own response potential. Operationally, this prescription specifies both the aggregate probability of the unresolved block and the reference zero of its effective free energy. When imposed across a family of response potentials defined for all finite system dimensions, this single-block replacement identity already implies conditional independence within the resolved block. The canonical Gibbs distribution, the Shannon entropy formula, and the Shannon grouping law then emerge as direct consequences, without requiring separate assumptions on conditional factorization or on the density of attainable probability vectors.

Our main quantitative results address what remains of this characterization when experiments are confined to a fixed energy window and a finite collection of system dimensions. On any bounded energy interval, three-state probability-ratio cycles control the additivity defect of binary log-odds within the same window. This yields a single endpoint-calibrated Gibbs scale, a dimension-independent bound in total-variation distance, and a discrete calibration protocol requiring only finitely many spectral preparations. To connect these log-odds bounds to errors in coarse-grained block probabilities, we introduce a finite suppression protocol that transfers control from an auxiliary configuration, in which the target block carries small probability mass, to the configuration of interest. Because probability residuals control only the outer gradient of the potential defect, recovering the internal conditional law involves a trade-off between the outer control displacement and a local bound on the second internal derivative, without requiring derivatives of the empirical measurement error.

Section 2 introduces response potentials and entropy duality. The exact characterization is proved in Section 3; finite-window estimates and their connection to coarse-graining errors follow in Sections 4 and 5. Section 6 examines reservoir models and shared temperature fluctuations. The appendices discuss finite-observation non-identifiability, sparse transition graphs, and numerical implementation.

2 Response potentials and entropy

Let $\Delta_N = \{p \in \mathbb{R}^N : p_i \geq 0, \sum_i p_i = 1\}$ denote the $(N - 1)$ -dimensional probability simplex and Δ_N° its relative interior. We consider diagonal state populations $P(\beta, E)$ parameterized by independently variable energy levels $E = (E_1, \dots, E_N) \in \mathbb{R}^N$ and inverse temperature $\beta > 0$. The entropy-integrability equations (EIE) read [1]

$$\partial_{E_j} P_i = \partial_{E_i} P_j, \quad (2.1)$$

$$\beta \partial_\beta P_i = \sum_j E_j \partial_{E_j} P_i. \quad (2.2)$$

Equation (2.1) expresses the closure of the reversible work differential $\sum_i P_i dE_i$ at fixed β , whereas (2.2) requires invariance under the reciprocal rescaling $(\beta, E) \mapsto (\lambda\beta, E/\lambda)$, which preserves the dimensionless spectrum $x = \beta E$. When a physical apparatus gives access only to a lower-dimensional submanifold of control parameters, (2.1)–(2.2), if satisfied on an ambient spectral domain, are sufficient for integrability on the accessible submanifold; integrability of the restricted response alone need not imply these ambient equations.

Proposition 2.1 (Local potential and physical normalization). *Any C^1 solution of (2.1)–(2.2) on a sufficiently small convex domain of dimensionless spectra admits the local representation $P(\beta, E) = \nabla \Psi(\beta E)$. Probability normalization $\sum_i P_i = 1$ is equivalent to translation covariance along the uniform direction $\mathbf{1} = (1, \dots, 1)^\top$,*

$$\Psi(x + t\mathbf{1}) = \Psi(x) + t, \quad (2.3)$$

whenever the segment connecting x and $x + t\mathbf{1}$ lies within the domain. The reversible work form is generated by the free energy $F(\beta, E) = \beta^{-1}\Psi(\beta E) + B(\beta)$. Requiring the thermodynamic internal energy $U = \partial_\beta(\beta F)$ to coincide with the spectral expectation $E \cdot P$ fixes $B(\beta) = C/\beta$ for a constant C . The associated thermodynamic entropy takes the Legendre form

$$S_{\text{th}} = x \cdot \nabla \Psi(x) - \Psi(x) + C_0, \quad dS_{\text{th}} = \beta \sum_i E_i dP_i. \quad (2.4)$$

Proof. The characteristic vector field $\beta\partial_\beta - E \cdot \nabla_E$ of (2.2) has coordinate invariants $x_i = \beta E_i$, so $P(\beta, E) = Q(x)$ locally. By (2.1), the differential 1-form $Q(x) \cdot dx$ is closed, and the Poincaré lemma on a convex domain yields $Q = \nabla \Psi$ for a scalar potential Ψ . Integrating the normalization identity $\mathbf{1} \cdot \nabla \Psi(x) = 1$ along $\mathbf{1}$ gives (2.3). Integrating $\partial_{E_i} F = P_i$ with respect to E produces $F(\beta, E) = \beta^{-1}\Psi(\beta E) + B(\beta)$. Differentiating βF with respect to β gives $\partial_\beta(\beta F) = E \cdot P + \frac{d}{d\beta}(\beta B(\beta))$, so imposing $U = E \cdot P$ requires $(\beta B)' = 0$. Finally, $d\Psi = P \cdot dx$ implies $d(x \cdot P - \Psi) = x \cdot dP$. Combined with the first-law identity $dU = P \cdot dE + E \cdot dP$, this recovers the Clausius relation (2.4) for the heat differential $\delta Q = E \cdot dP$. $\square$

In our global analysis, an N -state response potential is a finite C^1 function $\Psi_N : \mathbb{R}^N \rightarrow \mathbb{R}$ satisfying (2.3) with positive gradient $\nabla \Psi_N \in \Delta_N^\circ$. The potential is symmetric if it is invariant under coordinate permutations, and spectrally stable if it is concave on $\mathbb{R}^N$ (or equivalently on the zero-sum hyperplane $H_N = \mathbf{1}^\perp$). When Ψ_N is C^2 , translation covariance and concavity imply

$$\text{Hess } \Psi_N \mathbf{1} = 0, \quad v^\top \text{Hess } \Psi_N v \leq 0 \quad (\forall v \in \mathbb{R}^N), \quad C = -x^\top \text{Hess } \Psi_N(x) x \geq 0, \quad (2.5)$$

so that energy levels and populations respond monotonically in opposition ($\partial_{x_i} P_i \leq 0$) and the heat capacity C is nonnegative.

Proposition 2.2 (Concave entropy duality). *Let $\Psi : \mathbb{R}^N \rightarrow \mathbb{R}$ be a finite, concave, translation-covariant potential whose supergradients lie in Δ_N . Then the conjugate function*

$$s_\Psi(p) = \inf_{x \in \mathbb{R}^N} \{x \cdot p - \Psi(x)\}, \quad \Psi(x) = \min_{p \in \Delta_N} \{x \cdot p - s_\Psi(p)\}, \quad (2.6)$$

defines a proper, concave, upper semicontinuous entropy $s_\Psi : \Delta_N \rightarrow \mathbb{R} \cup \{-\infty\}$. Conversely, every proper, concave, upper semicontinuous function $s : \Delta_N \rightarrow \mathbb{R} \cup \{-\infty\}$ generates a finite, concave, translation-covariant potential Ψ via the second relation in (2.6), and at any differentiability point x of Ψ the unique minimizer is $p = \nabla \Psi(x)$. Moreover, if s_Ψ is finite throughout Δ_N° and Ψ is C^1 with $\nabla \Psi(x) \in \Delta_N^\circ$, then $\nabla \Psi$ is surjective onto the open simplex:

$$\nabla \Psi(\mathbb{R}^N) = \Delta_N^\circ. \quad (2.7)$$

Proof. As an infimum of affine functions, s_Ψ is concave and upper semicontinuous. Since Ψ is finite and concave on $\mathbb{R}^N$, it admits a supporting supergradient $p \in \Delta_N$ at every point x , which gives $s_\Psi(p) = x \cdot p - \Psi(x) > -\infty$; combining this with the definition of s_Ψ establishes the variational

recovery of $\Psi(x)$ in (2.6). Conversely, any proper upper semicontinuous concave function s on the compact simplex Δ_N is bounded above. For each x , the lower semicontinuous objective $p \mapsto x \cdot p - s(p)$ is bounded below and finite somewhere, so it attains a finite minimum in (2.6). This yields a finite, concave, translation-covariant potential Ψ whose supergradients coincide with the minimizers p . Differentiability of Ψ at x renders the supergradient unique. Finally, if s_Ψ is finite on Δ_N° , concave duality furnishes a supporting supergradient $x \in \mathbb{R}^N$ at every $p \in \Delta_N^\circ$, meaning p minimizes $x \cdot q - s_\Psi(q)$ and hence $p = \nabla \Psi(x)$. $\square$

Proposition 2.2 is the concave specialization of Fenchel–Moreau duality [2, 3] adapted to the simplex via translation covariance. Whenever $P = \nabla \Psi(x)$, the dual entropy $s_\Psi(P) = x \cdot P - \Psi(x)$ coincides with the Clausius entropy (2.4) up to the integration constant C_0 . For reference throughout the paper, we denote the canonical Gibbs response at scale parameter $a \in \mathbb{R}$ by

$$G_{a,N,i}(x) = \frac{e^{-ax_i}}{\sum_j e^{-ax_j}}. \quad (2.8)$$

Spectrally stable Gibbs responses correspond to $a \geq 0$, with $a = 0$ giving the uniform distribution. For $a > 0$, the canonically normalized potential and its dual entropy are

$$\Psi_{a,N}(x) = -a^{-1} \log \sum_j e^{-ax_j}, \quad s_{a,N}(p) = -a^{-1} \sum_i p_i \log p_i, \quad (2.9)$$

where a sets the inverse-temperature scale relative to β .

3 Exact replacement of an unresolved block

Suppose now that a symmetric, spectrally stable response potential Ψ_N is specified for every integer $N \geq 1$, normalized by $\Psi_1(t) = t$, and write $P_i^N = \partial_i \Psi_N$. Consider an n -state outer spectrum $x \in \mathbb{R}^n$ in which the a -th state ($1 \leq a \leq n$) is resolved into an internal block of m sublevels with relative spectrum $y \in \mathbb{R}^m$. The resolved $(n + m - 1)$ -state spectrum and the corresponding n -state coarse-grained spectrum are defined by

$$S_a(x, y) = (x_1, \dots, x_{a-1}, x_a + y_1, \dots, x_a + y_m, x_{a+1}, \dots, x_n), \quad (3.1)$$

$$T_a(x, y) = (x_1, \dots, x_{a-1}, x_a + \Psi_m(y), x_{a+1}, \dots, x_n). \quad (3.2)$$

To track the outer states in the resolved spectrum, let $\kappa_a(i) = i$ for $i < a$ and $\kappa_a(i) = i + m - 1$ for $i > a$.

Definition 3.1 (Single-block free-energy replacement). The family $\{\Psi_N\}_{N \geq 1}$ satisfies exact free-energy replacement if, for all $n, m \geq 1$, every $1 \leq a \leq n$, and all spectra $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$,

$$P_{\kappa_a(i)}^{n+m-1}(S_a(x, y)) = P_i^n(T_a(x, y)), \quad i \neq a, \quad (3.3)$$

$$\sum_{b=1}^m P_{a+b-1}^{n+m-1}(S_a(x, y)) = P_a^n(T_a(x, y)). \quad (3.4)$$

Although (3.4) follows from (3.3) by probability conservation, we state both relations explicitly because the aggregate block probability is a primary coarse-grained observable. Note that the value $\Psi_m(y)$ enters directly into the coarse-grained level $x_a + \Psi_m(y)$ in (3.2); consequently, shifting Ψ_m by an additive constant alters the replacement protocol even though it leaves the isolated m -state gradient $P^m(y) = \nabla \Psi_m(y)$ unchanged.

Theorem 3.2 (Exact free-energy replacement). *A family of symmetric, spectrally stable response potentials $\{\Psi_N\}_{N \geq 1}$ satisfies exact free-energy replacement if and only if there exists a universal constant $a_* > 0$ such that, for every $N \geq 1$,*

$$\Psi_N(x) = -a_*^{-1} \log \sum_{i=1}^N e^{-a_* x_i}. \quad (3.5)$$

In particular, conditional independence of internal block populations, the Shannon form of the dual entropy, and the Shannon grouping identity all follow as consequences of Definition 3.1.

Proof. Fix n, m, a and compare the resolved and coarse-grained potentials

$$F(x, y) = \Psi_{n+m-1}(S_a(x, y)), \quad G(x, y) = \Psi_n(T_a(x, y)).$$

By the chain rule, the replacement identities (3.3)–(3.4) are equivalent to $\nabla_x(F - G) = 0$, which implies $F(x, y) - G(x, y) = K(y)$ for a function K independent of x . Differentiating this relation with respect to y_b yields

$$\partial_{y_b} K(y) = P_{a+b-1}^{n+m-1}(S_a(x, y)) - P_a^n(T_a(x, y)) P_b^m(y). \quad (3.6)$$

For $n \geq 2$, choose an outer index $j \neq a$ and consider an auxiliary experiment in which state j of an n -state zero spectrum is split into k degenerate zero-energy sublevels. By permutation symmetry of the resulting $(n + k - 1)$ -state zero spectrum, every resolved state carries probability $(n + k - 1)^{-1}$. Applying (3.3) to this auxiliary split shows that the coarse-grained spectrum $t^{(k)} \in \mathbb{R}^n$ with $t_j^{(k)} = \Psi_k(0)$ and $t_i^{(k)} = 0$ ($i \neq j$) satisfies

$$P_a^n(t^{(k)}) = \frac{1}{n + k - 1} \xrightarrow{k \rightarrow \infty} 0.$$

Now, for any fixed $y \in \mathbb{R}^m$, choose $x^{(k)} \in \mathbb{R}^n$ by setting $x_i^{(k)} = t_i^{(k)}$ for $i \neq a$ and $x_a^{(k)} = t_a^{(k)} - \Psi_m(y)$, so that $T_a(x^{(k)}, y) = t^{(k)}$. Since both $P_{a+b-1}^{n+m-1}(S_a(x^{(k)}, y))$ and $P_a^n(t^{(k)}) P_b^m(y)$ lie in $[0, P_a^n(t^{(k)})]$, the right-hand side of (3.6) is bounded in absolute value by $(n + k - 1)^{-1}$. Because the left-hand side of (3.6) is independent of x , letting $k \rightarrow \infty$ forces $\partial_{y_b} K(y) = 0$, which establishes the conditional factorization

$$P_{a+b-1}^{n+m-1}(S_a(x, y)) = P_a^n(T_a(x, y)) P_b^m(y). \quad (3.7)$$

(For $n = 1$, (3.7) holds trivially by translation covariance.)

Specializing (3.7) to binary splits ($m = 2$) and invoking permutation symmetry to group any pair of states (i, j) of an N -state system ($N \geq 3$) into a two-state block shows that all pairwise log-odds depend only on the energy gap through the universal binary function h :

$$\log \frac{P_i^N(z)}{P_j^N(z)} = h(z_i - z_j), \quad h(t) = \log \frac{P_1^2(t, 0)}{P_2^2(t, 0)}.$$

Applying the identity $\frac{P_1^3}{P_3^3} = \frac{P_1^3}{P_2^3} \frac{P_2^3}{P_3^3}$ to a three-state spectrum gives Cauchy's functional equation $h(s+t) = h(s) + h(t)$ for all $s, t \in \mathbb{R}$. Since P^2 is continuous, $h(t) = -a_* t$ for a single constant $a_* \in \mathbb{R}$ common to all dimensions. Concavity of Ψ_2 implies that $P_1^2(t, 0)$ is nonincreasing in t , so $a_* \geq 0$. If $a_* = 0$, every response is uniform; taking $n = m = 2$ in (3.3) would then equate the resolved outer probability $1/3$ with the coarse-grained probability $1/2$, a contradiction. Hence $a_* > 0$.

Integrating $\nabla \Psi_m = G_{a_*, m}$ yields $\Psi_m(y) = -a_*^{-1} \log \sum_{b=1}^m e^{-a_* y_b} + k_m$ for a dimension-dependent constant k_m . Applying (3.3) to a two-state zero outer spectrum with one state split into m zero-energy sublevels gives

$$\frac{1}{m+1} = \frac{1}{1 + e^{-a_* \Psi_m(0)}},$$

so $\Psi_m(0) = -a_*^{-1} \log m$ and thus $k_m = 0$ for all m . The converse is immediate from the partition-function identity $\sum_{b=1}^m e^{-a_*(x_a + y_b)} = e^{-a_*(x_a + \Psi_m(y))}$. $\square$

Corollary 3.3 (Entropy and grouping). *Under the hypotheses of Theorem 3.2, the dual entropy on the closed simplex Δ_N is $s_N(p) = -a_*^{-1} \sum_{i=1}^N p_i \log p_i$ (with $0 \log 0 = 0$), and any partitioned distribution (p_{ab}) with block sums $q_a = \sum_b p_{ab}$ satisfies the Shannon grouping law*

$$s_{\sum_a n_a}((p_{ab})) = s_k(q) + \sum_{a:q_a>0} q_a s_{n_a}((p_{ab}/q_a)_b). \quad (3.8)$$

Proof. Evaluating the Fenchel conjugate (2.6) of Ψ_N in (3.5) yields $s_N(p) = -a_*^{-1} \sum_i p_i \log p_i$, and substituting $\log p_{ab} = \log q_a + \log(p_{ab}/q_a)$ for $q_a > 0$ directly produces (3.8). $\square$

Faddeev-type characterizations assume a recursive splitting rule on probability distributions [5]; Section 4 of Baez, Fritz, and Leinster [6] reviews this formulation. Theorem 3.2 derives both conditional independence and the grouping law from free-energy replacement in spectral space. The proof also identifies the role of unbounded auxiliary dimensions: splitting an external state into $k \rightarrow \infty$ degenerate sublevels provides a sequence of configurations in which the tested block probability $P_a^n(t^{(k)})$ becomes arbitrarily small.

3.1 Response independence versus potential additivity

For two subsystems with independent spectral controls $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$, the additive composite spectrum has levels $(x \oplus y)_{ij} = x_i + y_j$. Strict response independence requires the composite distribution to factorize across all spectral pairs:

$$P_{ij}^{NM}(x \oplus y) = P_i^N(x) P_j^M(y). \quad (3.9)$$

As a finite-spectrum counterpart to the decomposability theorems of Sandomirskiy and Tamuz [9, 10], strict response independence fixes the Gibbs family up to free-energy constants.

Proposition 3.4 (Strict product response independence). *Let $\{P^N\}_{N \geq 1}$ be a family of continuous, strictly positive, permutation-equivariant responses on all finite spectra. If (3.9) holds for all $N, M \geq 1$, then $P^N = \mathbf{G}_{a,N}$ for a universal constant $a \in \mathbb{R}$ (with $a \geq 0$ whenever P^N is generated by a concave potential). Unlike Definition 3.1, condition (3.9) leaves the additive constants of the generating potentials Ψ_N undetermined.*

Proof. Fix $x \in \mathbb{R}^N$ and distinct indices $i \neq j$, set $d = x_i - x_j$, and couple the system to a binary auxiliary spectrum $y = (0, d)$. Because the composite states $(i, 1)$ and $(j, 2)$ have identical total energy $x_i + 0 = x_j + d$, permutation symmetry of P^{2N} and (3.9) imply

$$P_i^N(x) P_1^2(0, d) = P_j^N(x) P_2^2(0, d).$$

Every pairwise probability ratio therefore depends solely on the energy gap via the binary ratio $P_2^2(0, d)/P_1^2(0, d)$, and the three-state cycle identity forces this dependence to be exponential as in the proof of Theorem 3.2. $\square$

By contrast, requiring the generating potentials to be additive under the product embedding,

$$\Psi_{NM}(x \oplus y) = \Psi_N(x) + \Psi_M(y) + C_{N,M}, \quad (3.10)$$

constrains only the directional derivatives along the subspace $\{x \oplus y\}$. Differentiating (3.10) with respect to x_i and y_j yields marginal consistency, $\sum_j P_{ij}^{NM}(x \oplus y) = P_i^N(x)$ and $\sum_i P_{ij}^{NM}(x \oplus y) = P_j^M(y)$, but does not enforce the joint factorization (3.9). Indeed, consider the two-scale canonical mixture

$$\Psi_N^\gamma(x) = \sum_{r \in \{r_1, r_2\}} \frac{\gamma_r}{r} \left(-\log \sum_{i=1}^N e^{-r x_i} \right), \quad 0 < r_1 < r_2, \quad \gamma_r > 0, \quad \gamma_{r_1} + \gamma_{r_2} = 1. \quad (3.11)$$

For every N , Ψ_N^γ is smooth, symmetric, translation covariant, and strictly concave on H_N , and satisfies $\min_i x_i - (\log N) \sum_r \gamma_r / r \leq \Psi_N^\gamma(x) \leq \min_i x_i$, so its dual entropy is finite throughout the

closed simplex Δ_N . Because each canonical summand is additive on $x \oplus y$, Ψ_N^γ satisfies (3.10) with $C_{N,M} = 0$. Yet for two identical nondegenerate binary systems with gap $z > 0$, setting $f_r = (1 + e^{rz})^{-1}$ gives

$$P_{11}^4(x \oplus x) - P_1^2(x)P_1^2(x) = \sum_r \gamma_r f_r^2 - \left(\sum_r \gamma_r f_r \right)^2 > 0,$$

violating (3.9). Since differentiability of Ψ_4^γ makes its supergradient unique, the Fenchel inequality at $z = x \oplus x$ is strict for the product distribution $p \otimes p$ with $p = P^2(x)$, so

$$s_4(p \otimes p) < z \cdot (p \otimes p) - \Psi_4^\gamma(z) = 2s_2(p). \quad (3.12)$$

Additivity of the response potential on product spectra is thus strictly weaker than statistical independence of the subsystem responses.

4 Finite-window Gibbs approximation

Theorem 3.2 assumes exact identities over all spectra and arbitrarily large auxiliary dimensions. Measurements instead give probabilities at finitely many spectra within a bounded energy window. We use two- and three-state calibrations of energy diameter at most $R > 0$, together with the conditional log-odds of the target system, to bound its total-variation distance from a common Gibbs law. The bound has no explicit dependence on the target dimension.

For $R > 0$, denote the domain of spectra with energy diameter at most R by

$$\mathcal{X}_N(R) = \left\{ x \in \mathbb{R}^N : \max_i x_i - \min_i x_i \leq R \right\}.$$

Let $P^N : \mathcal{X}_N(R) \rightarrow \Delta_N^\circ$ be permutation-equivariant and translation-invariant responses. We define the isolated binary log-odds function and its endpoint-calibrated inverse-temperature slope by

$$h(t) = \log \frac{P_1^2(t, 0)}{P_2^2(t, 0)}, \quad a_R = -\frac{h(R)}{R}. \quad (4.1)$$

Permutation symmetry ensures that $h(-t) = -h(t)$, and concavity of a generating binary potential Ψ_2 guarantees that $h(t)$ is nonincreasing with $a_R \geq 0$. At any spectrum $x \in \mathcal{X}_N(R)$, we measure the deviation of pairwise conditional probabilities from the isolated binary law via the conditional log-odds discrepancy

$$d_N(x) = \max_{i \neq j} \left| \log \frac{P_i^N(x)}{P_j^N(x)} - h(x_i - x_j) \right|, \quad (4.2)$$

which vanishes identically for $N = 2$.

To control the deviation of h from linearity on $[-R, R]$, we first record an elementary stability lemma for Cauchy's equation on a bounded interval (compare [12, 13]) together with a sharp conversion from pairwise log-odds bounds to total-variation distance.

Lemma 4.1 (Additivity on a bounded interval). *Let $h : [-R, R] \rightarrow \mathbb{R}$ be continuous and odd. If*

$$|h(s+t) - h(s) - h(t)| \leq \delta \quad \forall s, t \in [-R, R] \text{ with } s+t \in [-R, R],$$

then

$$\left| h(t) - \frac{h(R)}{R}t \right| \leq 2\delta, \quad |t| \leq R. \quad (4.3)$$

Proof. For $t \in [0, R]$, define the local average

$$A(t) = \frac{1}{R} \int_{-R}^0 (h(s+t) - h(s)) \, ds.$$

The approximate additivity hypothesis immediately implies $|A(t) - h(t)| \leq \delta$. Moreover, continuity of h allows us to differentiate $A(t)$ explicitly:

$$A'(t) = \frac{h(t) - h(t - R)}{R}.$$

Applying the additivity bound to the decomposition $(t - R) + R = t$ yields $|A'(t) - h(R)/R| \leq \delta/R$. Integrating from $t = 0$ (where $A(0) = 0$) gives $|A(t) - h(R)t/R| \leq (t/R)\delta \leq \delta$, and combining this with $|A(t) - h(t)| \leq \delta$ establishes (4.3) on $[0, R]$. The bound on $[-R, 0]$ follows by oddness. $\square$

Lemma 4.2 (Probability distance from pairwise log-odds). *Let $p, q \in \Delta_N^\circ$. If*

$$\max_{i,j} \left| \log \frac{p_i}{p_j} - \log \frac{q_i}{q_j} \right| \leq B,$$

then

$$\|p - q\|_1 \leq 2 \tanh(B/4), \quad \text{TV}(p, q) := \frac{1}{2} \|p - q\|_1 \leq \tanh(B/4). \quad (4.4)$$

Proof. Set $r_i = p_i/q_i$, $m = \min_i r_i$, and $M = \max_i r_i$, so that $m \leq 1 \leq M$ and $M/m \leq e^B$. If $m = M = 1$, the bound is trivial. Otherwise, convexity of $r \mapsto |r - 1|$ on $[m, M]$ together with $\sum_i q_i = 1$ and $\sum_i q_i r_i = 1$ yields

$$\text{TV}(p, q) = \frac{1}{2} \sum_{i=1}^N q_i |r_i - 1| \leq \frac{(M - 1)(1 - m)}{M - m}.$$

For a fixed ratio $r = M/m \geq 1$, substituting $M = rm$ shows that the right-hand side is maximized at $m = r^{-1/2}$, where it equals $(\sqrt{r} - 1)/(\sqrt{r} + 1) = \tanh((\log r)/4) \leq \tanh(B/4)$. $\square$

Theorem 4.3 (Finite-interval Gibbs estimate). *Suppose h is continuous on $[-R, R]$ and the three-state conditional discrepancy satisfies*

$$d_3(z) \leq \varepsilon_3 \quad \forall z \in \mathcal{X}_3(R).$$

Then for every $N \geq 2$ and every target spectrum $x \in \mathcal{X}_N(R)$ with $d_N(x) \leq \varepsilon_N$, setting $B_N = \varepsilon_N + 6\varepsilon_3$ yields

$$\left| \log \frac{P_i^N(x)}{P_j^N(x)} + a_R(x_i - x_j) \right| \leq B_N, \quad (4.5)$$

$$\|P^N(x) - \mathbf{G}_{a_R, N}(x)\|_1 \leq 2 \tanh(B_N/4). \quad (4.6)$$

Here the reference scale a_R is determined entirely by the binary response at gap R , with $a_R \geq 0$ when Ψ_2 is concave and $a_R > 0$ whenever $h(R) < 0$.

Proof. For any $s, t \in [-R, R]$ with $s + t \in [-R, R]$, the three-state spectrum $z = (s + t, t, 0)$ belongs to $\mathcal{X}_3(R)$. The pairwise log-odds of $P^3(z)$ sum to zero around the cycle $(1 \rightarrow 2 \rightarrow 3 \rightarrow 1)$. Approximating each log-ratio by h therefore gives

$$|h(s + t) - h(s) - h(t)| \leq 3\varepsilon_3.$$

Lemma 4.1 then gives $|h(u) + a_R u| \leq 6\varepsilon_3$ for all $u \in [-R, R]$. Applying the triangle inequality to (4.2) proves (4.5), and Lemma 4.2 yields (4.6). $\square$

Theorem 4.3 uses only spectra within the prescribed energy window. When $P^N = \nabla \Psi_N$ throughout $\mathcal{X}_N(R)$, integration along line segments from the origin also bounds the free-energy variation.

Corollary 4.4 (Free-energy variation). *Suppose $P^N = \nabla \Psi_N$ and the hypotheses of Theorem 4.3 hold uniformly for all $x \in \mathcal{X}_N(R)$ with error bound B_N . Define the centered reference potential*

$$\Phi_{a,N}(x) = \begin{cases} -a^{-1} \log(N^{-1} \sum_{i=1}^N e^{-ax_i}), & a \neq 0, \\ N^{-1} \sum_{i=1}^N x_i, & a = 0. \end{cases}$$

Then for every $x \in \mathcal{X}_N(R)$,

$$|\Psi_N(x) - \Psi_N(0) - \Phi_{a_R,N}(x)| \leq (\max_i x_i - \min_i x_i) \tanh(B_N/4). \quad (4.7)$$

Proof. For any $x \in \mathcal{X}_N(R)$, the ray tx ($0 \leq t \leq 1$) remains within $\mathcal{X}_N(R)$. Since any two probability vectors $p, q \in \Delta_N$ satisfy $|(p - q) \cdot x| \leq \text{TV}(p, q)(\max_i x_i - \min_i x_i)$, integrating $(\nabla \Psi_N(tx) - \mathbf{G}_{a_R,N}(tx)) \cdot x$ over $t \in [0, 1]$ and applying (4.6) yields (4.7). $\square$

4.1 Discrete calibration on a finite spectral grid

In an actual experiment or numerical simulation, one prepares only a finite number of spectra rather than the continuum $\mathcal{X}_3(R)$. We therefore formulate a discrete counterpart to Theorem 4.3 on a uniform calibration grid of mesh size $\Delta = R/M$ ($M \geq 2$).

Lemma 4.5 (Additivity on a finite grid). *Let $M \geq 2$ and let $h_{-M}, \dots, h_M \in \mathbb{R}$ satisfy $h_{-k} = -h_k$. If*

$$|h_{j+k} - h_j - h_k| \leq \delta \quad \forall j, k \geq 1 \text{ with } j + k \leq M,$$

then

$$\left| h_k - \frac{k}{M} h_M \right| \leq 2\delta \quad (-M \leq k \leq M). \quad (4.8)$$

Proof. By oddness, $|h_{j+k} - h_j - h_k| \leq \delta$ extends to all integers $j, k \in [-M, M]$ with $j + k \in [-M, M]$. For $0 \leq k \leq M$, define the discrete running average

$$A_k = \frac{1}{M} \sum_{s=-M}^{-1} (h_{s+k} - h_s).$$

Then $A_0 = 0$, $|A_k - h_k| \leq \delta$, and telescoping the sum gives the increment

$$A_k - A_{k-1} = \frac{h_{k-1} - h_{k-1-M}}{M}.$$

Applying approximate additivity to the decomposition $(k - 1 - M) + M = k - 1$ shows that $|(A_k - A_{k-1}) - h_M/M| \leq \delta/M$. Summing over $1 \leq j \leq k$ yields $|A_k - (k/M)h_M| \leq (k/M)\delta \leq \delta$, which combines with $|A_k - h_k| \leq \delta$ to give (4.8). $\square$

The prefactor 2 in (4.8) is sharp uniformly in M under endpoint calibration: the odd sequence defined by $h_0 = h_M = 0$ and $h_k = 3k/M - 1$ for $1 \leq k < M$ has additivity defect 1 on the positive grid while attaining $h_{M-1} = 2 - 3/M \rightarrow 2$ as $M \rightarrow \infty$. Subtracting the linear trend $3k/M$ renders the sequence nonincreasing without altering either its additivity defect or its deviation from the endpoint-calibrated line, showing that monotonicity does not reduce this worst-case constant.

To calibrate the response and test a finite collection of target spectra $\mathcal{T}$, we prepare:

1. the M binary spectra $(0, k\Delta)$ for $1 \leq k \leq M$;
2. the $\binom{M}{2}$ ternary spectra $(0, j\Delta, k\Delta)$ for $1 \leq j < k \leq M$;
3. each target spectrum $x \in \mathcal{T}$ (with $\max_i x_i - \min_i x_i \leq R$).

The total number of distinct spectral preparations is at most

$$\mathcal{E} = M + \binom{M}{2} + |\mathcal{T}|. \quad (4.9)$$

Let $h_k = h(k\Delta)$ for $1 \leq k \leq M$, with $h_0 = 0$ and $h_{-k} = -h_k$, and let $\varepsilon_3^{\text{grid}}$ denote the maximum conditional log-odds discrepancy across the $\binom{M}{2}$ ternary grid preparations. For any $t \in [-R, R]$, let $\kappa(t) \in \{-M, \dots, M\}$ be the integer nearest to t/Δ (breaking ties oddly), and define the discrete target discrepancy

$$d_{N,M}^{\text{grid}}(x) = \max_{i \neq j} \left| \log \frac{P_i^N(x)}{P_j^N(x)} - h_{\kappa(x_i - x_j)} \right|,$$

which compares each pairwise log-odds in the target spectrum against the nearest calibrated binary gap.

Theorem 4.6 (Gibbs estimate from finitely many preparations). *Suppose the binary response is generated by a differentiable concave potential ($a_R \geq 0$). For every target spectrum $x \in \mathcal{T} \cap \mathcal{X}_N(R)$, setting*

$$B_{N,M}(x) = d_{N,M}^{\text{grid}}(x) + 6\varepsilon_3^{\text{grid}} + \frac{a_R \Delta}{2}$$

yields

$$\|P^N(x) - \mathbf{G}_{a_R, N}(x)\|_1 \leq 2 \tanh(B_{N,M}(x)/4). \quad (4.10)$$

If all pairwise energy gaps of x lie on the calibration grid $\Delta\mathbb{Z}$, the rounding term $a_R \Delta/2$ drops out of $B_{N,M}(x)$.

Proof. For any $j, k \geq 1$ with $j + k \leq M$, the ternary preparation $(0, j\Delta, (j+k)\Delta)$ yields $|h_{j+k} - h_j - h_k| \leq 3\varepsilon_3^{\text{grid}}$, so Lemma 4.5 gives $|h_m + a_R m \Delta| \leq 6\varepsilon_3^{\text{grid}}$ for all $-M \leq m \leq M$. Since $|\kappa(x_i - x_j)\Delta - (x_i - x_j)| \leq \Delta/2$ and $a_R \geq 0$, the triangle inequality bounds the pairwise log-odds discrepancy relative to $\mathbf{G}_{a_R, N}(x)$ by $B_{N,M}(x)$, and Lemma 4.2 completes the proof. $\square$

Theorem 4.6 bounds the error at the tested spectra $x \in \mathcal{T}$. A Lipschitz bound on the target log-odds and a finite ℓ^∞ -net extend this estimate to the full window $\mathcal{X}_N(R)$.

Corollary 4.7 (Uniform extension from a finite spectral net). *Fix $N \geq 2$ and suppose $\mathcal{T}_N \subset \mathcal{X}_N(R)$ is a finite ν -net up to translations, in the sense that for every $x \in \mathcal{X}_N(R)$ there exist $z \in \mathcal{T}_N$ and $c \in \mathbb{R}$ with $\|x - z - c\mathbf{1}\|_\infty \leq \nu$. Assume further that the log-odds functions $\ell_{ij}(x) = \log(P_i^N(x)/P_j^N(x))$ are L_ℓ -Lipschitz with respect to $\|\cdot\|_\infty$ on $\mathcal{X}_N(R)$ (for instance, $L_\ell \leq 2K/m$ whenever $P_i^N \geq m > 0$ and $\|\nabla P_i^N\|_1 \leq K$). If $|\ell_{ij}(z) + a_R(z_i - z_j)| \leq B_*$ holds for all $z \in \mathcal{T}_N$, then every $x \in \mathcal{X}_N(R)$ satisfies*

$$\|P^N(x) - \mathbf{G}_{a_R, N}(x)\|_1 \leq 2 \tanh\left(\frac{B_* + (L_\ell + 2|a_R|)\nu}{4}\right). \quad (4.11)$$

In particular, fixing each coordinate in turn at zero and letting the remaining $N-1$ coordinates range over $\{0, \Delta_0, \dots, R\}$ with $\Delta_0 = R/M_0$ constructs such a net of cardinality at most $N(M_0 + 1)^{N-1}$ with mesh radius $\nu = \Delta_0/2$.

Proof. By translation invariance, choosing $z \in \mathcal{T}_N$ and $c \in \mathbb{R}$ with $\|x - z - c\mathbf{1}\|_\infty \leq \nu$ gives $|\ell_{ij}(x) - \ell_{ij}(z)| \leq L_\ell \nu$ and $|(x_i - x_j) - (z_i - z_j)| \leq 2\nu$. Combining these inequalities with the net bound B_* and applying Lemma 4.2 yields (4.11). $\square$

4.2 Finite-sample concentration guarantees

When the response probabilities of each of the $\mathcal{E}$ preparations are estimated from n independent experimental trials, the discrete protocol admits an explicit finite-sample confidence bound. Let V denote the total number of outcome coordinates summed across all $\mathcal{E}$ preparations. For a prescribed failure probability $\zeta \in (0, 1)$, set

$$t_n = \sqrt{\frac{\log(2V/\zeta)}{2n}},$$

and let $\hat{p}$ denote the empirical frequency vectors, with $\hat{m}$ the minimum empirical probability observed across all V coordinates. Whenever $\hat{m} > t_n$, define

$$\Lambda_n = \frac{2t_n}{\hat{m} - t_n}.$$

From the empirical frequencies, form the plug-in estimators $\hat{h}_k = \log(\hat{p}_2/\hat{p}_1)$ on the binary preparations $(0, k\Delta)$ (with $\hat{h}_0 = 0$ and $\hat{h}_{-k} = -\hat{h}_k$), the empirical ternary discrepancy $\hat{\epsilon}_3^{\text{grid}}$, the empirical target discrepancy $\hat{d}_{N,M}^{\text{grid}}(x)$, and the empirical slope

$$\hat{a} = -\frac{\hat{h}_M}{R}, \quad \hat{a}_+ = \max\{0, \hat{a}\}.$$

Corollary 4.8 (Finite-sample guarantee). *Under the assumptions of Theorem 4.6, suppose each preparation is sampled n times independently. With probability at least $1 - \zeta$, on the event $\hat{m} > t_n$, every target spectrum $x \in \mathcal{T} \cap \mathcal{X}_N(R)$ satisfies*

$$\|P^N(x) - \mathbf{G}_{\hat{a}_+, N}(x)\|_1 \leq 2 \tanh(\hat{B}_{N,M}(x)/4), \quad (4.12)$$

$$\hat{B}_{N,M}(x) = \hat{d}_{N,M}^{\text{grid}}(x) + 6\hat{\epsilon}_3^{\text{grid}} + \frac{\hat{a}_+ \Delta}{2} + 2\Lambda_n, \quad (4.13)$$

where $2\Lambda_n$ may be tightened to Λ_n whenever $\hat{a} \geq 0$, and the rounding term $\hat{a}_+ \Delta/2$ vanishes if all pairwise gaps of x lie on $\Delta\mathbb{Z}$.

Proof. By Hoeffding's inequality [14] and a union bound over all V outcome coordinates, the event $\max |\hat{p}_i - p_i| \leq t_n$ holds with probability at least $1 - \zeta$. On this event, whenever $\hat{m} > t_n$, the elementary bound $|\log(\hat{p}_i/\hat{p}_j) - \log(p_i/p_j)| \leq 2t_n/(\hat{m} - t_n) = \Lambda_n$ holds for every empirical log-odds. Because the empirical ternary frequencies satisfy the three-state ratio cycle identically, applying Lemma 4.5 directly to the sequence $\hat{h}_k$ yields $|\hat{h}_k + \hat{a}k\Delta| \leq 6\hat{\epsilon}_3^{\text{grid}}$ without incurring an extra sampling penalty at the algebraic step. Passing from the empirical target log-odds to the true target log-odds adds at most Λ_n . Finally, if $\hat{a} < 0$, concavity of Ψ_2 implies $h(R) \leq 0$, so $|\hat{a}_+ - \hat{a}|R = |\hat{h}_M| \leq |\hat{h}_M - h(R)| \leq \Lambda_n$; clipping the slope to $\hat{a}_+ = 0$ therefore adds at most one further Λ_n . Applying Lemma 4.2 completes the proof. $\square$

The condition $\hat{m} > t_n$ is sufficient for the Hoeffding confidence intervals used here to have positive lower endpoints, allowing a finite uniform log-odds error bound. It is not a necessary threshold for other confidence procedures; rare outcomes nevertheless make log-odds estimation less precise.

4.3 Numerical illustrations

To illustrate the discrete calibration protocol of Theorem 4.6, we take $R = 3$ and $M = 24$, corresponding to 24 binary and 276 ternary calibration preparations, and evaluate four target spectra of dimensions 3, 4, 6, 8 (for $\mathcal{E} = 304$ preparations in total). Expressed in units of R , the target spectra are $(0, 0.243, 1)$, $(0, 0.25, 0.5, 1)$, $(0, 0.04, 0.07, 0.35, 0.877, 1)$, and $(j/7)_{j=0}^7$. We compare four response families: the exact Gibbs response at unit scale ($a = 1$), equal-weight two-scale mixtures with scales $(0.85, 1.15)$ and $(0.5, 1.5)$, and a three-scale mixture $(0.4, 1, 2)$ with weights $(0.25, 0.5, 0.25)$. The resulting endpoint-calibrated slopes a_R are 1, 0.970434, 0.744790, and 0.804275, respectively. Across the four target spectra, the maximum true total-variation distances from $\mathbf{G}_{a_R, N}$ are 0, 0.007390, 0.064792, and 0.068035, while the corresponding upper bounds from Theorem 4.6 are 0.029367, 0.105157, 0.575179, and 0.603586 (Figure 1, left). For the exact Gibbs model, the on-grid four-state target $(0, 0.25, 0.5, 1)R$ exhibits zero bound and zero error. The maximum off-grid bound 0.029367, attained at the six-state target, combines the rounded-target discrepancy $d_{6,24}^{\text{grid}} = 0.055$ and the explicit allowance $a_R \Delta/2 = 0.0625$, giving $\tanh((0.055 + 0.0625)/4)$ despite zero actual error.

Figure 2 illustrates the finite-sample bound of Corollary 4.8 for the weak two-scale mixture $(0.85, 1.15)$ with $R = 2$ and $M = 8$ ($\mathcal{E} = 40$ preparations and $V = 121$ outcome coordinates), using

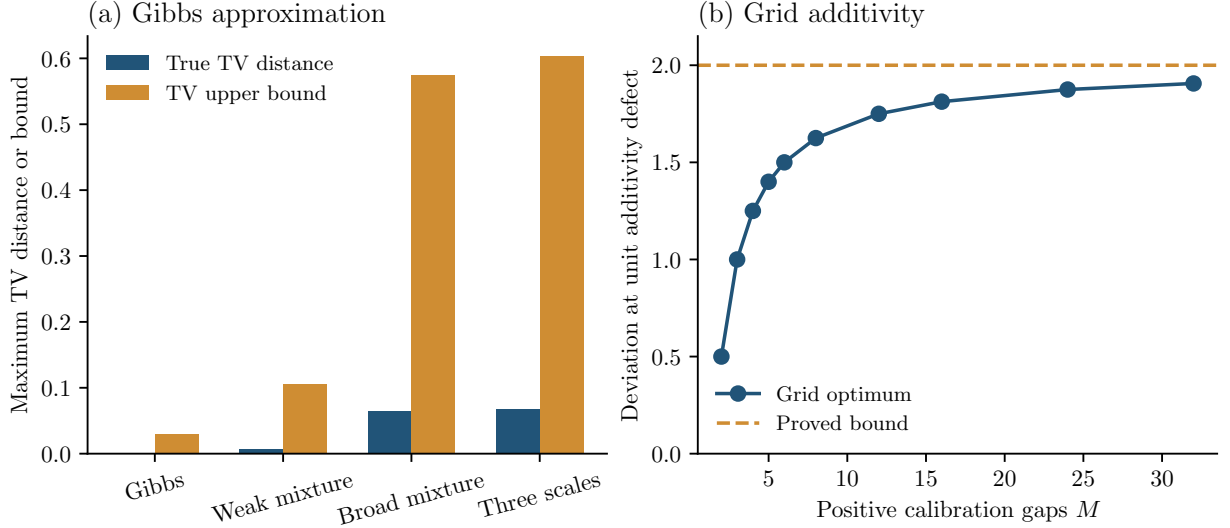


Fig. 1 Finite-preparation Gibbs bounds and sharpness of the grid additivity constant. Left: largest true total-variation distance and the corresponding upper bound from Theorem 4.6 across the four target spectra for each response family. Right: worst-case deviation from the endpoint-calibrated line over odd sequences with unit additivity defect, computed by linear programming; at $M = 32$ the maximum is 1.90625, approaching the sharp asymptotic constant 2 of Lemma 4.5.

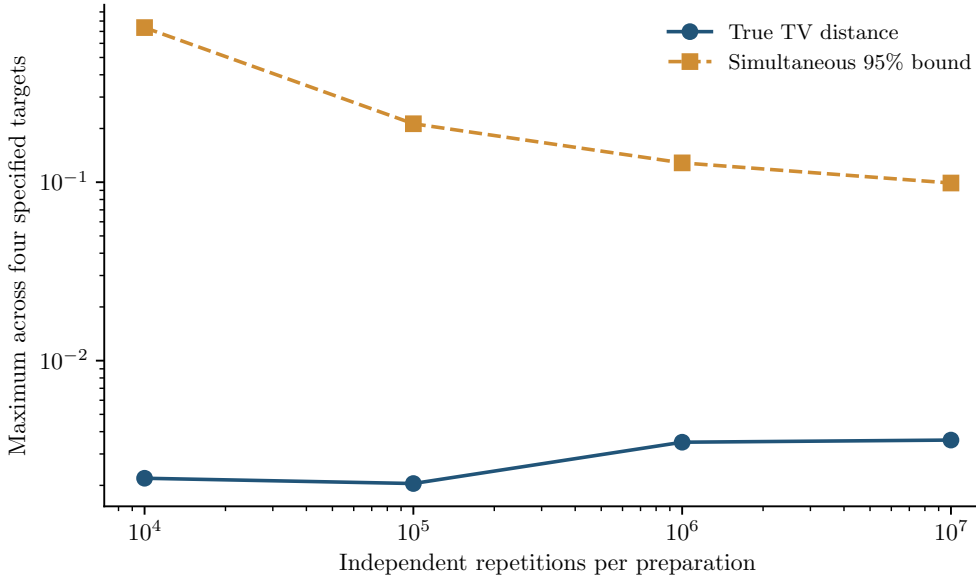


Fig. 2 Finite-sample Gibbs approximation bounds for the equal mixture of scales 0.85 and 1.15 ($R = 2$, $M = 8$). Curves display maxima across the four target spectra of the true total-variation distance to $\mathcal{G}_{\hat{a}_+, N}$ and the 95% confidence bound of Corollary 4.8 as functions of the number of repetitions n per preparation.

multinomial samples of size $n \in \{10^4, 10^5, 10^6, 10^7\}$ per preparation at confidence level $1 - \zeta = 0.95$. As n increases, the sampling penalty Λ_n decays as $n^{-1/2}$ and the empirical upper bound converges toward the deterministic finite-grid bound.

5 Coarse-graining error bounds

The stability theorems of Section 4 take conditional log-odds discrepancies $d_N(x)$ as their input. We now establish how $d_N(x)$ can itself be bounded from coarse-grained block-probability residuals using only a finite family of symmetric, spectrally stable C^2 response potentials.

Consider a binary split ($m = 2$) of outer state a in an n -state system, producing a resolved system of dimension $N = n + 1$. The potential defect between the resolved and coarse-grained descriptions is

$$K(x, y) = \Psi_{n+1}(S_a(x, y)) - \Psi_n(T_a(x, y)), \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^2, \quad (5.1)$$

and the observable coarse-graining probability residuals are

$$r_i(x, y) = \begin{cases} P_{\kappa_a(i)}^{n+1}(S_a(x, y)) - P_i^n(T_a(x, y)), & i \neq a, \\ \sum_{b=1}^2 P_{a+b-1}^{n+1}(S_a(x, y)) - P_a^n(T_a(x, y)), & i = a. \end{cases} \quad (5.2)$$

By the chain rule, $r_i(x, y) = \partial_{x_i} K(x, y)$. Thus, experimental bounds on the probability residuals r_i directly control the outer gradient $\nabla_x K$, whereas the conditional distribution inside the split block is governed by the internal derivative $\partial_{y_1} K$. The following theorem bridges these two derivatives by combining an outer control displacement to a block-suppressing configuration x^* with a finite-difference estimate in the internal coordinate y_1 .

Theorem 5.1 (Finite-control conditional estimate). *Fix a target configuration $(x, y) \in \mathbb{R}^n \times \mathbb{R}^2$, a suppressing outer field $x^* \in \mathbb{R}^n$ at ℓ^1 distance $D = \|x - x^*\|_1$, and an internal half-width $h_0 > 0$. Assume that:*

- (i) *the probability residuals satisfy $|r_i| \leq e$ for all $1 \leq i \leq n$ throughout the two-dimensional control rectangle*

$$\mathcal{C} = \{(x + s(x^* - x), y + te_1) : 0 \leq s \leq 1, |t| \leq h_0\}; \quad (5.3)$$

- (ii) *along the two boundary lines $z \in \{x, x^*\}$ and $|t| \leq h_0$, the internal second derivative is bounded by*

$$\left| \frac{d^2}{dt^2} K(z, y + te_1) \right| \leq L; \quad (5.4)$$

- (iii) *at the suppressing field x^* , the coarse-grained block probability satisfies $u^* = P_a^n(T_a(x^*, y)) \leq \eta$;*
- (iv) *at the target configuration (x, y) , the resolved block mass M and the internal conditional probabilities q, v satisfy the conditioning bounds*

$$M = \sum_{b=1}^2 P_{a+b-1}^{n+1}(S_a(x, y)) \geq \mu > 0, \quad (5.5)$$

$$q = \frac{P_a^{n+1}(S_a(x, y))}{M}, \quad v = P_1^2(y), \quad q, v \in [\rho, 1 - \rho],$$

for some $\rho \in (0, 1/2]$.

Then, for every step size $h \in (0, h_0]$,

$$\left| \log \frac{P_a^{n+1}(S_a(x, y))}{P_{a+1}^{n+1}(S_a(x, y))} - \log \frac{P_1^2(y)}{P_2^2(y)} \right| \leq \frac{\eta + 2e + De/h + Lh}{\mu\rho(1 - \rho)}. \quad (5.6)$$

In particular, whenever $De > 0$, $L > 0$, and $\sqrt{De/L} \leq h_0$, optimizing over h at $h = \sqrt{De/L}$ yields the numerator

$$\eta + 2e + 2\sqrt{DLe}. \quad (5.7)$$

Proof. Define the outer difference $E(t) = K(x, y + te_1) - K(x^*, y + te_1)$ for $|t| \leq h_0$. Integrating $\nabla_x K = r$ along the segment connecting x^* to x gives $|E(t)| \leq De$ for all $|t| \leq h_0$, while assumption (ii) gives $|E''(t)| \leq 2L$. Taylor's theorem with integral remainder around $t = 0$ then bounds the central difference:

$$\left| E'(0) - \frac{E(h) - E(-h)}{2h} \right| \leq \frac{1}{2h} \int_{-h}^h |E'(0) - E'(t)| dt \leq Lh, \quad |E'(0)| \leq \frac{De}{h} + Lh.$$

At the suppressing configuration x^* , let $p^* = P_a^{n+1}(S_a(x^*, y))$. Because the resolved block mass at x^* differs from u^* by at most e , we have $0 \leq p^* \leq u^* + e$. Differentiating K with respect to y_1 at (x^*, y) gives $\partial_{y_1} K(x^*, y) = p^* - u^*v$, whence $|\partial_{y_1} K(x^*, y)| \leq \eta + e$. Combining this with $|E'(0)| = |\partial_{y_1} K(x, y) - \partial_{y_1} K(x^*, y)|$ yields

$$|\partial_{y_1} K(x, y)| \leq \eta + e + \frac{De}{h} + Lh.$$

At the target configuration (x, y) , writing $u = P_a^n(T_a(x, y))$, we have $\partial_{y_1} K(x, y) = Mq - uv$ and $|M - u| \leq e$. Consequently,

$$|q - v| = \frac{|Mq - uv + (u - M)v|}{M} \leq \frac{\eta + 2e + De/h + Lh}{M} \leq \frac{\eta + 2e + De/h + Lh}{\mu}.$$

Since the derivative of the logit map $w \mapsto \log(w/(1-w))$ on $[\rho, 1-\rho]$ is bounded by $[\rho(1-\rho)]^{-1}$, the mean-value theorem gives (5.6). $\square$

The square-root scaling $\sqrt{DLe}$ in (5.7) reflects the ill-posedness of numerical differentiation when only zero-order probability residuals r_i are measured. By the chain rule, $\frac{d^2}{dt^2} \Psi_n(T_a(z, y + te_1)) = \partial_{aa} \Psi_n(T_a)(P_1^2)^2 + P_a^n(T_a) \partial_{11} \Psi_2$, so if the diagonal Hessian entries of $\Psi_{n+1}, \Psi_n, \Psi_2$ are bounded in magnitude by H_{n+1}, H_n, H_2 , the internal curvature constant may be chosen as

$$L = H_{n+1} + H_n + H_2. \quad (5.8)$$

To realize the suppression bound $u^* \leq \eta$ constructively using a finite auxiliary system, we employ a finite-dimensional version of the degenerate refinement used in the proof of Theorem 3.2.

Proposition 5.2 (Suppression via finite degenerate refinement). *Fix $n \geq 2$, an outer index $j \neq a$, and an integer $k \geq 2$. If replacing k degenerate zero-energy states in an $(n+k-1)$ -state zero spectrum incurs an outer probability residual of at most e_0 , then the coarse-grained spectrum $t^* \in \mathbb{R}^n$ with $t_j^* = \Psi_k(0)$ and $t_i^* = 0$ ($i \neq j$) satisfies*

$$P_a^n(t^*) \leq \frac{1}{n+k-1} + e_0. \quad (5.9)$$

Consequently, for any internal binary spectrum $y \in \mathbb{R}^2$, setting $x_i^* = t_i^*$ ($i \neq a$) and $x_a^* = t_a^* - \Psi_2(y)$ ensures $T_a(x^*, y) = t^*$, so that condition (iii) of Theorem 5.1 holds with $\eta = (n+k-1)^{-1} + e_0$.

Proof. By permutation symmetry, every state of the $(n+k-1)$ -state zero spectrum has probability $(n+k-1)^{-1}$, and the replacement residual bounds $|P_a^n(t^*) - (n+k-1)^{-1}| \leq e_0$. $\square$

Proposition 5.2 makes explicit the finite resource trade-off underlying block suppression: reducing the residual mass $(n+k-1)^{-1}$ requires preparing an auxiliary system of $n+k-1$ states and setting the j -th outer level to $\Psi_k(0)$. This control requirement enters through $D = \|x - x^*\|_1$ and may increase as k grows. Combining Theorem 5.1 with the finite-window theorems of Section 4 yields a complete chain from coarse-graining residuals to Gibbs approximation.

Corollary 5.3 (Gibbs approximation from coarse-graining residuals). *Fix an energy window $R > 0$ and a finite set of dimensions containing 2 and 3. Suppose every pair of levels in each tested spectrum $x \in \mathcal{X}_N(R)$ ($N \geq 3$) can be represented as a binary split satisfying Theorem 5.1 with dimension-dependent bounds $\eta_N, e_N, D_N, L_N, h_N, \mu_N, \rho_N$, and set*

$$\theta_N = \inf_{0 < h \leq h_N} \frac{\eta_N + 2e_N + D_N e_N / h + L_N h}{\mu_N \rho_N (1 - \rho_N)}, \quad \theta_2 = 0. \quad (5.10)$$

If these bounds hold across $\mathcal{X}_3(R)$, then every tested $x \in \mathcal{X}_N(R)$ satisfies

$$\|P^N(x) - G_{a_R, N}(x)\|_1 \leq 2 \tanh\left(\frac{\theta_N + 6\theta_3}{4}\right), \quad a_R = -\frac{1}{R} \log \frac{P_1^2(R, 0)}{P_2^2(R, 0)} \geq 0. \quad (5.11)$$

Similarly, if the three-state tests are performed only on the $\binom{M}{2}$ ternary grid preparations of Theorem 4.6, then (5.11) holds for every tested target spectrum whose pairwise gaps lie on $\Delta\mathbb{Z}$.

Proof. By translation invariance, the isolated binary log-odds of a split pair with internal spectrum y equals $h(y_1 - y_2)$, so Theorem 5.1 implies $d_N(x) \leq \theta_N$. Applying Theorem 4.3 in the continuum case, or Theorem 4.6 in the discrete grid case, yields (5.11). $\square$

When Proposition 5.2 is used with $e_0 \leq e_N$, the numerator of (5.10) becomes

$$\frac{1}{N+k-2} + 3e_N + \frac{D_N e_N}{h} + L_N h, \quad (5.12)$$

For each fixed k , taking $e_N \rightarrow 0$ with the control and conditioning bounds fixed, and then $h \rightarrow 0$, leaves only the suppression contribution. Letting $k \rightarrow \infty$ then removes this contribution, provided the target conditioning bounds remain positive. This recovers exact conditional factorization; extending the hypotheses to all finite spectra gives the characterization of Theorem 3.2.

5.1 Discretizing the control rectangle

If the probability residuals r_i in condition (i) of Theorem 5.1 are evaluated on a discrete grid over the parameter rectangle $(s, t) \in [0, 1] \times [-h_0, h_0]$ with mesh spacings (Δ_s, Δ_t) (including the boundary lines) and satisfy the Lipschitz modulus

$$|r_i(s, t) - r_i(s', t')| \leq A_s |s - s'| + A_t |t - t'|, \quad (5.13)$$

then an empirical maximum residual $\hat{e}$ measured within tolerance σ_r extends to the continuum rectangle with bound

$$e = \hat{e} + \sigma_r + \frac{1}{2} A_s \Delta_s + \frac{1}{2} A_t \Delta_t. \quad (5.14)$$

When each residual is obtained as the difference of two directly measured block probabilities with individual error at most σ , we have $\sigma_r \leq 2\sigma$. The present finite-difference and interpolation estimates use the a priori bounds (5.4) and (5.13), respectively. Appendix A shows that finite observations alone do not uniquely determine the global response; it does not assert that these particular regularity assumptions are necessary for every quantitative interpolation method.

5.2 Analytic bounds for two-scale mixtures

To illustrate the progression from coarse-graining residuals to conditional log-odds bounds and finally to finite-grid total-variation bounds, we analyze the symmetric two-scale canonical mixture with scales $r_{\pm} = 1 \pm \delta$ ($0 < \delta < 1$):

$$\Psi_N^{(\delta)}(z) = \frac{1}{2} \left[\frac{\Psi_N^{\text{can}}(r_- z)}{r_-} + \frac{\Psi_N^{\text{can}}(r_+ z)}{r_+} \right], \quad \Psi_N^{\text{can}}(z) = -\log \sum_{i=1}^N e^{-z_i}. \quad (5.15)$$

Here every coarse-grained spectrum $T_a(z, y)$ retains the exact potential value $\Psi_2^{(\delta)}(y)$ defined in (5.15).

Proposition 5.4 (Analytic probability-error envelope). *For the two-scale family (5.15), the response deviation from the unit-scale Gibbs distribution satisfies*

$$\|P^{N,(\delta)}(z) - \mathbf{G}_{1,N}(z)\|_1 \leq \frac{\delta^2}{4} \text{osc}(z)^2, \quad \text{osc}(z) = \max_i z_i - \min_i z_i. \quad (5.16)$$

On any binary control rectangle (5.3), let R_f denote the maximum oscillation of the resolved spectrum S_a across its four vertices, and set

$$R_c = \max_{z \in \{x, x^*\}} \text{osc}(T_a(z, y)) + h_0, \quad R_y = \text{osc}(y) + h_0, \quad b_\Psi = \delta^2 \left[\frac{R_y^2}{8r_-} + \frac{\log 2}{r_-^3} \right]. \quad (5.17)$$

Then every coarse-graining probability residual on $\mathcal{C}$ is bounded uniformly by

$$e = \frac{\delta^2}{8}(R_f^2 + R_c^2) + \frac{b_\Psi}{4}, \quad (5.18)$$

and the internal curvature constant in Theorem 5.1 may be taken as $L = 3/4$.

Proof. For the canonical family $p_i(r) = e^{-rz_i} / \sum_j e^{-rz_j}$, direct differentiation gives $p_i''(r) = p_i(r)[(z_i - \langle z \rangle_r)^2 - \text{Var}_r(z)]$, so $\sum_i |p_i''(r)| \leq 2 \text{Var}_r(z) \leq \text{osc}(z)^2/2$ by Popoviciu's inequality. Taylor expansion of the symmetric average $\frac{1}{2}[p(1-\delta) + p(1+\delta)]$ around $r = 1$ yields (5.16), and half of this ℓ^1 bound controls the probability discrepancy of any single event.

Similarly, for a binary spectrum w , differentiating $\phi_w(r) = r^{-1}\Psi_2^{\text{can}}(rw)$ twice gives $\phi_w''(r) = -r^{-1} \text{Var}_r(w) - 2r^{-3}H(p(r))$, where $H(p) = -\sum_{i=1}^2 p_i \log p_i \leq \log 2$. Since $\text{Var}_r(w) \leq \text{osc}(w)^2/4$, symmetric Taylor expansion on $[r_-, r_+]$ bounds $|\Psi_2^{(\delta)}(w) - \Psi_2^{\text{can}}(w)| \leq b_\Psi$ whenever $\text{osc}(w) \leq R_y$.

By convexity of $\text{osc}(\cdot)$, the resolved oscillation on $\mathcal{C}$ is maximized at the four vertices, giving $\text{osc}(S_a) \leq R_f$, while $|\partial_{y_1} \Psi_2^{(\delta)}| \leq 1$ implies $\text{osc}(T_a) \leq R_c$ on $\mathcal{C}$. Decomposing each residual r_i by inserting the unit-scale canonical resolved and coarse-grained probabilities (which cancel identically when evaluated at the canonical block potential Ψ_2^{can}) bounds $|r_i|$ by the sum of the resolved mixture error $\delta^2 R_f^2/8$, the coarse mixture error $\delta^2 R_c^2/8$, and the shift in canonical coarse probability induced by $|\Psi_2^{(\delta)} - \Psi_2^{\text{can}}| \leq b_\Psi$. Because $|\partial_j \mathbf{G}_{1,n,i}| \leq 1/4$, the last contribution is at most $b_\Psi/4$, establishing (5.18). Finally, each diagonal Hessian entry of $\Psi_m^{(\delta)}$ is bounded in magnitude by $(r_- + r_+)/8 = 1/4$, so (5.8) yields $L = 3/4$. $\square$

We apply Proposition 5.4 first to the target spectrum $(0, 1/2, 1)$, represented as a binary split of its first two levels with $x = (0, 1)$ and $y = (0, 1/2)$, using the suppressing field $x^* = (12, 1)$ ($D = 12$) and internal half-width $h_0 = 1/2$. The step size is chosen as $h = \min\{h_0, \sqrt{De/L}\}$. To obtain a full finite-grid total-variation bound via Corollary 5.3, we take $R = 1$ and $M = 4$, giving the six ternary grid spectra $\mathcal{T}_3 = \{(0, j/4, k/4) : 1 \leq j < k \leq 4\}$. Assigning each of the three pairs in every ternary spectrum its own control rectangle with $D = 12$, $h_0 = 1/2$, and $L = 3/4$, and using the analytical lower bounds

$$\mu_3 = \frac{2}{2 + e^{(1+\delta)R}}, \quad \rho_3 = \frac{1}{1 + e^{(1+\delta)R}},$$

yields the uniform grid bound $\max_{z \in \mathcal{T}_3} \text{TV}(P^{3,(\delta)}(z), \mathbf{G}_{a_R,3}(z)) \leq \tanh(7\theta_3/4)$ shown in Figure 3. The true conditional and total-variation discrepancies are $O(\delta^2)$ as $\delta \rightarrow 0$. The square-root contribution $\sqrt{DL}e$ in Theorem 5.1 instead decreases linearly with δ when it dominates the bound. For smaller δ , the fixed suppression error η determines the limiting bound.

6 Reservoir models and temperature fluctuations

We now examine how concrete physical reservoir models and temperature-fluctuation mechanisms behave under the two integrability conditions (2.1) and (2.2). Setting $k_B = 1$ and $x = \beta E$, we quantify violations of work-form closure on x -dependent responses via the antisymmetric defect

$$\mathcal{R}_{ij}(x) = \partial_{x_j} P_{N,i}(x) - \partial_{x_i} P_{N,j}(x),$$

and violations of temperature-energy scaling via the Euler scaling operator

$$D = \beta \partial_\beta - \sum_{i=1}^N E_i \partial_{E_i}.$$

The conditions $\mathcal{R} = 0$ and $DP_N = 0$ probe distinct physical properties: the former tests whether state populations are generated by a reversible work potential, whereas the latter tests whether E_i and β^{-1} are the sole energy scales governing the reduced state.

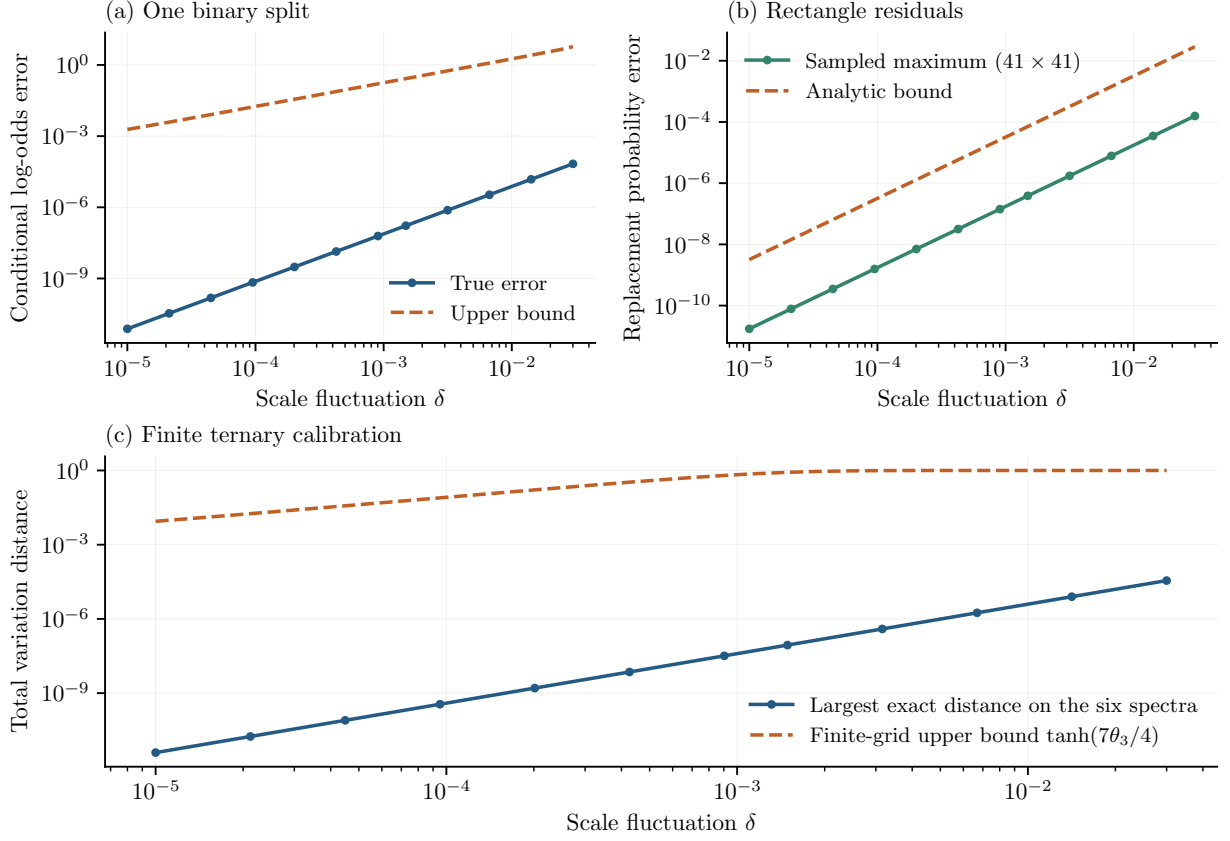


Fig. 3 Coarse-graining error propagation for the symmetric two-scale mixture $r_{\pm} = 1 \pm \delta$ with control displacement $D = 12$ and internal half-width $h_0 = 1/2$. (a) True conditional log-odds discrepancy for the binary split of $(0, 1/2, 1)$ and the theoretical upper bound of Theorem 5.1. (b) Analytic continuum residual envelope e from Proposition 5.4 compared against the maximum residual sampled on a 41×41 grid over $\mathcal{C}$. (c) Maximum total-variation distance across the six ternary grid spectra ($R = 1, M = 4$) and the resulting bound $\tanh(7\theta_3/4)$ from Corollary 5.3.

6.1 Microcanonical reservoir with constant heat capacity

Consider a weakly coupled finite thermal reservoir whose density of states $\omega_B(u)$ defines the entropy $S_B(u) = \log[\omega_B(u)/\omega_B(u_0)] = C \log(u/u_0)$ for $u > 0$. In this convention, the temperature is $T_B = (S'_B)^{-1} = u/C$ and the heat capacity is $C > 0$. For the same density of states, the phase-space-volume entropy used by Campisi et al. [15] gives heat capacity $C + 1$, explaining the corresponding unit shift in the power-law exponent. Calibrating the total microcanonical energy $a = C/\beta$ so that the baseline reservoir inverse temperature is $S'_B(a) = \beta$ independent of the system spectrum E produces the weights

$$P_{N,i}^{(C)}(x) = \frac{(1 - x_i/C)^C}{\sum_j (1 - x_j/C)^C}, \quad x_i < C. \quad (6.1)$$

Proposition 6.1 (Closure defect for a fixed-energy-calibrated reservoir). *For $N \geq 2$, the finite-reservoir response (6.1) satisfies $DP_N^{(C)} = 0$, but its closure defect for $i \neq j$ is*

$$\mathcal{R}_{ij}^{(C)}(x) = \frac{P_{N,i}^{(C)}(x)P_{N,j}^{(C)}(x)(x_j - x_i)}{C(1 - x_i/C)(1 - x_j/C)}. \quad (6.2)$$

Consequently, $P_N^{(C)}$ is not generated by a response potential on any open domain containing nondegenerate spectra. As $C \rightarrow \infty$, writing $p_i(x) = e^{-x_i} / \sum_j e^{-x_j}$, one has uniformly on compact subsets of $\mathbb{R}^N$:

$$P_{N,i}^{(C)}(x) = p_i(x) \left[1 - \frac{x_i^2 - \sum_j p_j(x)x_j^2}{2C} \right] + O(C^{-2}), \quad (6.3)$$

$$\mathcal{R}_{ij}^{(C)}(x) = \frac{p_i(x)p_j(x)(x_j - x_i)}{C} + O(C^{-2}). \quad (6.4)$$

Proof. Because $P_N^{(C)}$ depends on (β, E) solely through $x = \beta E$, $DP_N^{(C)} = 0$ holds identically. Writing $\phi_C(t) = C \log(1 - t/C)$ with $\phi'_C(t) = -(1 - t/C)^{-1}$, differentiation gives $\partial_{x_j} P_{N,i}^{(C)} = P_{N,i}^{(C)} P_{N,j}^{(C)} (1 - x_j/C)^{-1}$ for $i \neq j$, and antisymmetrizing yields (6.2). Expanding $\phi_C(t) = -t - t^2/(2C) + O(C^{-2})$ uniformly on compact sets yields (6.3)–(6.4). $\square$

Physically, the failure of closure in (6.2) stems from holding the total energy $a = C/\beta$ fixed while shifting the system spectrum: a uniform shift $E_i \mapsto E_i + d$ reduces the residual reservoir energy $a - E_i$ unless a is shifted simultaneously by d , violating translation invariance. Notice also that for any reservoir with positive heat capacity ($S_B''(a) = -\beta_B^2/C_B < 0$), the log-weight $\log w_B(E) = S_B(a - E)$ is strictly concave in E , whereas for any nondegenerate positive superposition of Boltzmann factors $B(E) = \int_0^\infty e^{-bE} \nu(db)$,

$$\frac{d^2}{dE^2} \log B(E) = \text{Var}_{\nu_E}(b) \geq 0, \quad \nu_E(db) = \frac{e^{-bE} \nu(db)}{B(E)}.$$

The opposite signs of the second derivatives preclude representing finite-bath microcanonical weights as positive temperature mixtures (a distinction also noted in explicit finite Ising-chain calculations [16]).

6.2 Spectrum-dependent reservoir calibration

If instead the total energy calibration is adjusted with the arithmetic mean of the system energy levels so as to preserve translation invariance, a finite microcanonical bath can generate an exact noncanonical EIE potential.

Proposition 6.2 (Spectrum-calibrated non-Gibbs potential). *For a reservoir with entropy $S_B(u) = \log(u/u_0)$ ($C = 1$), calibrating the total energy as*

$$a(\beta, E) = \bar{E} + \beta^{-1}, \quad \bar{E} = \frac{1}{N} \sum_{i=1}^N E_i,$$

yields on the domain $\max_i(x_i - \bar{x}) < 1$ the affine response

$$P_{N,i}(x) = \frac{1 - (x_i - \bar{x})}{N}, \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (6.5)$$

which satisfies both EIE conditions (2.1)–(2.2) and is generated by the following quadratic potential, strictly concave on H_N :

$$\Psi_N(x) = \bar{x} - \frac{1}{2N} \sum_{i=1}^N (x_i - \bar{x})^2, \quad \nabla^2 \Psi_N = -\frac{1}{N} \left(I - \frac{1}{N} \mathbf{1}\mathbf{1}^\top \right). \quad (6.6)$$

Proof. Since $e^{S_B(a-E_i)} \propto a - E_i = \beta^{-1}[1 - (x_i - \bar{x})]$ and $\sum_{i=1}^N [1 - (x_i - \bar{x})] = N$, normalizing gives (6.5), which is the gradient of (6.6). $\square$

Because $a(\beta, E + d\mathbf{1}) = a(\beta, E) + d$, this calibration keeps the reservoir excitations $a - E_i$ invariant under uniform shifts of the energy zero. The quadratic entropy $s_N(p) = -\frac{N}{2} \sum_{i=1}^N (p_i - 1/N)^2$ on Δ_N generates a global concave extension of the local potential (6.6) through Fenchel duality. Its gradient is the Euclidean projection of $N^{-1}(\mathbf{1} - x)$ onto Δ_N . It agrees with (6.5) on the stated positivity domain; outside that domain, some populations can vanish, so the extension need not belong to the strictly positive global response class used in Section 3.

6.3 Hamiltonians of mean force at fixed coupling

For a system coupled to a canonical bath at inverse temperature β with total Hamiltonian $H_{\text{tot}} = H_S + H_B + V$, the Hamiltonian of mean force H^* is defined by [17, 18]

$$e^{-\beta H^*} = \frac{\text{Tr}_B e^{-\beta H_{\text{tot}}}}{Z_B}, \quad Z_B = \text{Tr}_B e^{-\beta H_B}.$$

When the bare system levels E_i are varied while holding the bath Hamiltonian H_B and interaction V fixed, the bare populations automatically satisfy the closure condition (2.1), even for noncommuting quantum interactions.

Proposition 6.3 (Closure at fixed system–bath coupling). *Let $H_{\text{tot}}(E) = \sum_{i=1}^N E_i \Pi_i + H_B + V$, where the orthogonal projectors Π_i sum to the identity and, together with the bath Hamiltonian H_B and coupling V , are independent of E . Assume the partition functions are finite and the trace derivatives below are justified, as is automatic in finite dimension. Then the mean-force free energy $F^*(\beta, E) = -\beta^{-1} \log(Z_{\text{tot}}/Z_B)$ generates the bare populations $P_{N,i} = \text{Tr}(\rho_{\text{tot}} \Pi_i)$:*

$$\partial_{E_i} F^* = P_{N,i}, \quad \partial_{E_j} P_{N,i} = \partial_{E_i} P_{N,j}. \quad (6.7)$$

Proof. By Duhamel’s formula for the derivative of the matrix exponential and cyclicity of the trace, $\partial_{E_i} Z_{\text{tot}} = -\beta \text{Tr}(\Pi_i e^{-\beta H_{\text{tot}}})$. Since Z_B does not depend on E , differentiating F^* gives $\partial_{E_i} F^* = P_{N,i}$, and symmetry of mixed partial derivatives yields (6.7). $\square$

Despite satisfying (6.7), fixed-coupling models generally violate the scaling equation (2.2) because H_B and V introduce fixed energy scales that are not rescaled with E . As a concrete solvable example, consider a classical two-state bath $b \in \{0, 1\}$ with joint energies $H(i, b) = E_i + (\Delta + g_i)b$, where $\Delta > 0$ and $h_i = \Delta + g_i > 0$. Writing $n_i = (1 + e^{\beta h_i})^{-1}$ for the conditional bath excitation probability, we obtain

$$P_{N,i}(\beta, E) = \frac{e^{-\beta E_i} (1 + e^{-\beta h_i})}{\sum_j e^{-\beta E_j} (1 + e^{-\beta h_j})}, \quad (6.8)$$

$$D P_{N,i} = -\beta P_{N,i} \left(h_i n_i - \sum_{j=1}^N P_{N,j} h_j n_j \right), \quad (6.9)$$

$$D \log \frac{P_{N,i}}{P_{N,j}} = -\beta (h_i n_i - h_j n_j). \quad (6.10)$$

At a fixed temperature, $D P_N = 0$ precisely when all quantities $h_i n_i$ coincide. Distinct state-dependent couplings therefore produce a scaling defect generically, although accidental cancellations can occur at particular temperatures. Furthermore, the thermodynamic internal energy conjugate to F^* ,

$$U^* = \partial_\beta(\beta F^*) = \langle H^* + \beta \partial_\beta H^* \rangle = \sum_{i=1}^N P_{N,i} (E_i + h_i n_i) - \frac{\Delta}{1 + e^{\beta \Delta}}, \quad (6.11)$$

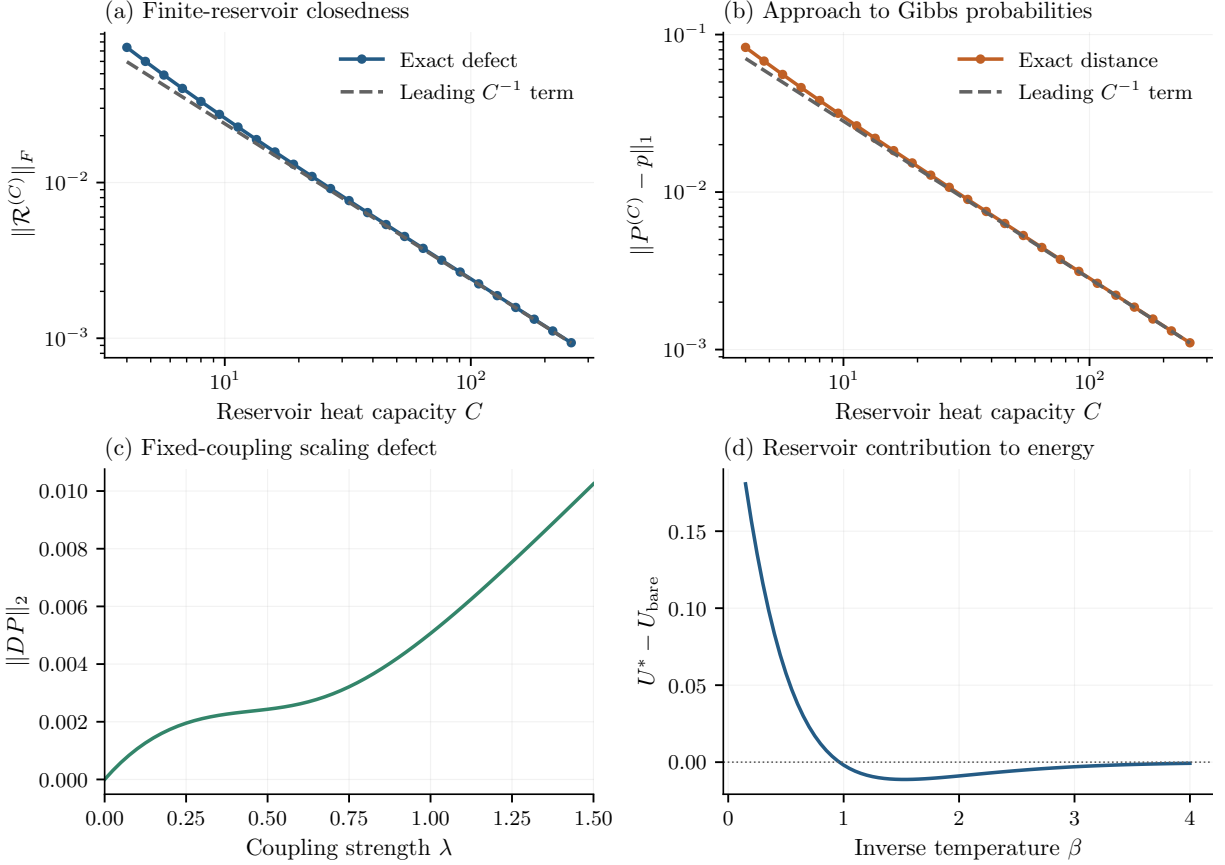


Fig. 4 Complementary integrability defects for finite-capacity and fixed-coupling reservoirs on the spectrum $E = (0, 0.7, 1.8)$. (a) Frobenius norm $\|\mathcal{R}^{(C)}\|_F$ of the closure defect and (b) population distance $\|P^{(C)} - \mathbb{G}_{1,N}\|_1$ at $\beta = 1$ for the constant-capacity reservoir (6.1), compared with the leading C^{-1} asymptotics (6.3)–(6.4) (dashed lines). (c) Euclidean norm $\|DP_N\|_2$ of the scaling defect at $\beta = 1$, $\Delta = 1$, and $g = \lambda(0, 0.4, 1.2)$, and (d) difference between the mean-force internal energy U^* and bare energy U_{bare} as β varies at $\lambda = 1$ for the fixed-coupling model (6.8), whose work form is closed.

differs from the bare spectral expectation $U_{\text{bare}} = \sum_i P_{N,i} E_i$ by the shift in bath and interaction energy. Figure 4 contrasts the closure defect of the constant-capacity microcanonical reservoir against the scaling defect and energy renormalization of this fixed-coupling model.

6.4 Shared temperature fluctuations and mutual information

Suppose now that a system experiences slow fluctuations in its effective inverse temperature, described by a probability measure γ on $(0, \infty)$ independent of (β, E) [1, 19]:

$$p_{N,i}(x) = \frac{e^{-x_i}}{\sum_j e^{-x_j}}, \quad P_{N,i}^\gamma(x) = \int_{(0,\infty)} p_{N,i}(rx) \gamma(dr).$$

By subtracting the zero-field value before integrating, the generating potential can be defined for an arbitrary probability measure γ without requiring $\int r^{-1} \gamma(dr) < \infty$.

Proposition 6.4 (Centered mixture potential). *For every probability measure γ on $(0, \infty)$, the integral*

$$\tilde{\Psi}_N^\gamma(x) = \int_{(0,\infty)} \frac{\Psi_N^{\text{can}}(rx) - \Psi_N^{\text{can}}(0)}{r} \gamma(dr), \quad \Psi_N^{\text{can}}(x) = -\log \sum_{i=1}^N e^{-x_i}, \quad (6.12)$$

defines a finite, concave, C^1 , translation-covariant response potential with gradient $\nabla \tilde{\Psi}_N^\gamma = P_N^\gamma$. If $\int_{(0,\infty)} r \gamma(dr) < \infty$, then $\tilde{\Psi}_N^\gamma$ is C^2 with negative semidefinite Hessian

$$\nabla^2 \tilde{\Psi}_N^\gamma(x) = \int_{(0,\infty)} r (p_N(rx) p_N(rx)^\top - \text{diag } p_N(rx)) \gamma(dr) \preceq 0.$$

Proof. Since $\min_i x_i \leq r^{-1} [\Psi_N^{\text{can}}(rx) - \Psi_N^{\text{can}}(0)] \leq \max_i x_i$ and $p_{N,i}(rx) \in [0, 1]$ uniformly in $r > 0$, dominated convergence justifies differentiation under the integral sign. The Hessian integrand is bounded in operator norm by $2r$, so a finite first moment guarantees twice continuous differentiability. $\square$

When two noninteracting subsystems A and B with spectra $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$ are exposed to the same fluctuating temperature scale $R \sim \gamma$, their joint distribution

$$P_{ij}^{AB} = \int_{(0,\infty)} p_{N,i}(rx) p_{M,j}(ry) \gamma(dr)$$

becomes statistically correlated: for two identical binary systems with gap $z \neq 0$, $P_{11}^{AB} - P_1^A P_1^B = \text{Var}_\gamma((1 + e^{Rz})^{-1})$, which vanishes if and only if γ is a Dirac mass. For weak fluctuations around unit scale, the resulting mutual information is governed universally by the product of the canonical heat capacities of the two subsystems.

Proposition 6.5 (Mutual information from shared temperature fluctuations). *Let $R = e^{\tau\xi}$, where ξ is a bounded, symmetrically distributed random variable with $\mathbb{E}[\xi^2] = 1$. For fixed spectra $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$, let $p_i = p_{N,i}(x)$ and $q_j = p_{M,j}(y)$, and denote their reference canonical heat capacities by*

$$C_A = \sum_{i=1}^N p_i \left(x_i - \sum_k p_k x_k \right)^2, \quad C_B = \sum_{j=1}^M q_j \left(y_j - \sum_k q_k y_k \right)^2.$$

Then, as $\tau \rightarrow 0$, the mutual information between A and B expands as

$$I(A : B) = \frac{\tau^4}{2} C_A C_B + O(\tau^6), \quad (6.13)$$

with remainder uniform for (x, y) in compact sets.

Proof. Set $L_x = x \cdot \nabla_x$ and $L_y = y \cdot \nabla_y$. Because ξ is symmetric, all odd moments vanish and Taylor expansion around $\tau = 0$ yields the covariance matrix

$$\delta_{ij} := P_{ij}^{AB} - P_i^A P_j^B = \tau^2 (L_x p_i)(L_y q_j) + O(\tau^4).$$

Since $\sum_i \delta_{ij} = \sum_j \delta_{ij} = 0$, expanding the relative entropy $I(A : B) = \sum_{i,j} P_{ij}^{AB} \log(P_{ij}^{AB}/(P_i^A P_j^B))$ to second order in δ_{ij} and using $P_i^A = p_i + O(\tau^2)$, $P_j^B = q_j + O(\tau^2)$ gives

$$I(A : B) = \frac{1}{2} \sum_{i,j} \frac{\delta_{ij}^2}{P_i^A P_j^B} + O(\tau^6) = \frac{\tau^4}{2} \left(\sum_{i=1}^N \frac{(L_x p_i)^2}{p_i} \right) \left(\sum_{j=1}^M \frac{(L_y q_j)^2}{q_j} \right) + O(\tau^6).$$

Substituting $L_x p_i = -p_i(x_i - \sum_k p_k x_k)$ identifies the two factors as C_A and C_B . $\square$

If the symmetry assumption on ξ is relaxed to $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[\xi^2] = 1$, with boundedness retained, the leading term $\frac{1}{2} \tau^4 C_A C_B$ remains unchanged with remainder $O(\tau^5)$. Figure 5 illustrates the leading law and the predicted sixth-order remainder for the symmetric two-point distribution $\Pr(\xi = \pm 1) = 1/2$.

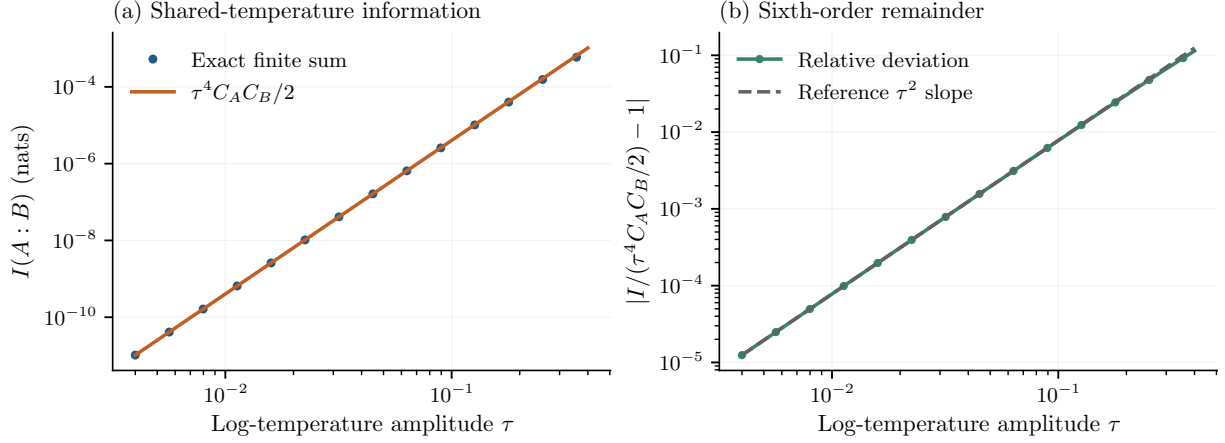


Fig. 5 Subsystem mutual information generated by a shared binary temperature fluctuation $R = e^{\pm\tau}$ with equal weights, for spectra $x = (0, 0.8, 1.9)$ and $y = (0, 1.1)$. (a) Exact mutual information $I(A : B)$ and the leading heat-capacity product law $\frac{1}{2}\tau^4 C_A C_B$ from Proposition 6.5. (b) Absolute relative deviation $|I(A : B)/(\frac{1}{2}\tau^4 C_A C_B) - 1|$, illustrating the $O(\tau^2)$ relative error and hence the $O(\tau^6)$ absolute remainder in (6.13).

6.5 Boundary entropy and low-temperature asymptotics

Spectral stability and concave duality do not by themselves imply finite entropy at pure ground states ($\beta \rightarrow \infty$). For example, the Cauchy response on a two-level system with gap $\Delta > 0$ and $z = \beta\Delta$,

$$f(z) = \frac{1}{2} - \frac{1}{\pi} \arctan z,$$

has Clausius entropy $S(z) = zf(z) - \int_0^z f(t) dt = -\frac{1}{2\pi} \log(1 + z^2)$ and heat capacity $C(z) = \frac{z^2}{\pi(1+z^2)} \rightarrow \frac{1}{\pi}$ as $\beta \rightarrow \infty$. Equivalently, its dual entropy on Δ_2 ,

$$s_2(p) = \frac{1}{\pi} \log \sin(\pi p),$$

diverges to $-\infty$ at the pure-state boundaries $p \in \{0, 1\}$. By contrast, finite mixtures of positive canonical scales (3.11) possess bounded dual entropies throughout the closed simplex Δ_N and vanishing heat capacity as $\beta \rightarrow \infty$, while still violating the Shannon grouping law.

7 Discussion

Our analysis establishes a direct quantitative link between operational coarse-graining and the canonical Gibbs distribution in finite systems. At the exact level, requiring a block of energy levels to be replaceable by its own response potential across all finite spectra forces conditional independence within the block and uniquely selects the Gibbs family with its canonical partition-function normalization. At the quantitative level, restricting observations to a bounded energy window of diameter R shows that two- and three-state preparations suffice to calibrate a common inverse-temperature scale a_R , while pairwise conditional log-odds discrepancies control the total-variation distance to the Gibbs law uniformly in the target dimension N .

Theorem 5.1 bounds conditional log-odds discrepancies from coarse-grained block-probability residuals and a finite block-suppression experiment. The estimate separates the residual block mass η , the probability error e , the outer control displacement D , and the internal curvature bound L . Its denominator records the conditioning of the target probabilities through μ and ρ . Measurement and interpolation tolerances contribute to e . Increasing the auxiliary degeneracy k reduces the suppression term to $(N + k - 2)^{-1}$, at the cost of preparing more states and controlling the level shift $\Psi_k(0)$.

Work-form closure and temperature-energy scaling distinguish the reservoir models in Section 6. A finite-capacity microcanonical bath with fixed total-energy calibration satisfies scaling but violates closure at order C^{-1} . A fixed-coupling mean-force model has a closed work form with respect to the bare levels, but its bath and interaction energies generally break scaling. The smooth temperature-scale mixtures considered here satisfy both integrability equations. A shared fluctuating temperature nevertheless correlates noninteracting subsystems: for weak bounded symmetric log-temperature fluctuations, the leading mutual information is $\frac{1}{2}\tau^4 C_A C_B$. Integrability of each subsystem therefore remains compatible with a measurable failure of joint statistical independence.

A Non-identifiability from finitely many observations

The following construction shows that exact agreement with the canonical Gibbs family at finitely many spectra, including finitely many derivatives and coarse-graining replacement tests, does not uniquely determine a smooth stable response potential globally. This non-uniqueness is compatible with quantitative bounds between observations obtained from concavity or additional regularity assumptions, such as those used in Corollary 4.7 and Section 5.1.

Proposition A.1 (Indistinguishability under finite observations). *Fix any finite collection of potential evaluations and finitely many derivative evaluations at finitely many spectra across finitely many dimensions (including all inner, coarse, and resolved spectra of finitely many replacement experiments), and suppose all recorded values agree with the unit-scale canonical Gibbs family $\Psi_{\text{can},N}(x) = -\log \sum_{i=1}^N e^{-x_i}$. Then for any chosen dimension $N_0 \geq 2$, there exists a smooth, symmetric, spectrally stable, strictly positive response family $\{\tilde{\Psi}_N\}_{N \geq 1}$ that matches every recorded observation identically, has a finite dual entropy on each closed simplex Δ_N , and yet differs from the Gibbs family in dimension N_0 .*

Proof. Project all observed N_0 -dimensional spectra onto the zero-sum hyperplane $H_{N_0} = \mathbf{1}^\perp$ and take the union of their coordinate-permutation orbits; this forms a finite subset $\mathcal{O} \subset H_{N_0}$. Choose a smooth, nontrivial, permutation-invariant bump function $b : H_{N_0} \rightarrow \mathbb{R}$ with compact support disjoint from a neighborhood of $\mathcal{O}$, and define

$$\tilde{\Psi}_{N_0}(x) = \Psi_{\text{can},N_0}(x) + \epsilon b(x - \bar{x}\mathbf{1}), \quad \bar{x} = \frac{1}{N_0} \sum_{i=1}^{N_0} x_i,$$

while keeping $\tilde{\Psi}_N = \Psi_{\text{can},N}$ for all $N \neq N_0$. Translation covariance and permutation symmetry hold by construction. Because $\text{supp}(b)$ is compact, $\nabla \Psi_{\text{can},N_0}$ is bounded away from the boundary of Δ_{N_0} and $\text{Hess } \Psi_{\text{can},N_0}$ is uniformly negative definite on H_{N_0} over $\text{supp}(b)$. For sufficiently small $\epsilon \neq 0$, $\tilde{\Psi}_{N_0}$ therefore retains a strictly positive gradient and a negative definite Hessian on H_{N_0} . Since b vanishes in a neighborhood of $\mathcal{O}$, all recorded functional and derivative values—as well as all coarse-grained spectra constructed from them—coincide with their canonical values. The perturbation has a nonzero gradient somewhere, so the response differs from the unit-scale Gibbs law. Since it agrees with that law outside the compact projected support, it cannot be a Gibbs response at any other constant scale. Finally, since $|\tilde{\Psi}_{N_0} - \Psi_{\text{can},N_0}|$ is uniformly bounded, (2.6) ensures that $s_{\tilde{\Psi}_{N_0}}$ differs from the Shannon entropy by a bounded amount and hence remains finite on Δ_{N_0} . $\square$

B Pairwise probability ratios and sparse transition graphs

The pairwise log-odds condition used in Section 3 requires the ratio $P_i^N(x)/P_j^N(x)$ between every pair of states to depend solely on the energy gap $x_i - x_j$. In a continuous-time Markov jump process with transition rates $W_{i \leftarrow j} > 0$ obeying detailed balance $P_j W_{i \leftarrow j} = P_i W_{j \leftarrow i}$, this all-pairs condition holds whenever every pair of states is connected by a transition whose forward-to-backward rate ratio is a universal function of the energy difference, $W_{i \leftarrow j}/W_{j \leftarrow i} = R(x_i - x_j)$.

By contrast, imposing gap-local rate ratios only along the edges of a sparse transition graph does not enforce the Gibbs distribution at a fixed spectrum. Consider, for example, the three-state linear

chain $1 \leftrightarrow 2 \leftrightarrow 3$ with fixed spectrum $x = (0, 1, 3)$ and edge rate ratios $W_{i \leftarrow j}/W_{j \leftarrow i} = e^{-(x_i - x_j)^3}$ on the two allowed transitions $(1, 2)$ and $(2, 3)$. The unique stationary distribution satisfies detailed balance with weights

$$P \propto (1, e^{-1}, e^{-9}),$$

since $(x_2 - x_1)^3 = 1$ and $(x_3 - x_2)^3 = 8$, so that no single inverse-temperature scale a reproduces both population ratios $P_2/P_1 = e^{-1}$ and $P_3/P_1 = e^{-9}$. Complete connectivity alone at a fixed spectrum does not exclude non-exponential laws. If detailed balance with the same continuous gap-ratio function is required for every independently variable three-state spectrum, the cycle identity instead forces $R(u + v) = R(u)R(v)$ and hence $R(u) = e^{-au}$. On a chain, the same conclusion follows if the endpoint ratio is independent of the intermediate level for every choice of both endpoint energies. A separate sufficient condition is complete passivity, meaning passivity of every finite tensor power under arbitrary joint cyclic operations [20]. For a strictly positive state of a finite system with a nonconstant Hamiltonian, complete passivity selects a Gibbs state at nonnegative inverse temperature. If the Hamiltonian is proportional to the identity, every state is completely passive and no such selection follows.

C Numerical implementation and reproducibility

Online Resource 1 implements the finite-window calibration protocol (Section 4), the coarse-graining control-rectangle bounds (Section 5), and the reservoir and fluctuating-temperature models (Section 6). Spectra, mixture weights, grid resolutions, and sample sizes are specified in the text and in the programs. The multinomial simulations use seed 26052026; the random probability-distance checks use seed 27042026. Fixed seeds make sampling repeatable within a fixed software environment, whose versions are recorded in the supplied summaries. Floating-point results may differ slightly across platforms. The programs compare analytical derivatives with finite differences, test the scaling expression (6.9), measure the decay of the mutual-information remainder, compute the finite-grid linear-programming optima, and check equality cases of Lemma 4.2. These numerical checks accompany the proofs and do not establish the continuum regularity assumptions used in the error bounds.

Statements and Declarations

Funding

Sun Zhenyu acknowledges support from Northeastern University under Grant No. 10011345052601.

Data and code availability

Online Resource 1 ([ESM_1.zip](#)) contains the Python source code, generated CSV data tables, figure-generation scripts, and numerical verification summaries supporting this study. Its [README.md](#) provides instructions for reproducing all numerical results and figures.

Competing interests

The authors declare no competing interests.

Use of AI-assisted tools

ChatGPT (OpenAI) was used for proofreading and language editing, mathematical consistency checks, assistance with local corrections to statements and proofs, and preparation and verification of the LaTeX and computational materials. The authors are responsible for the mathematical arguments, numerical results, and final text.

References

- [1] Zhai, R.-X., Sun, C.P.: Quantum thermodynamic integrability for canonical and noncanonical statistics. *Physical Review Letters* **134**, 160404 (2025) <https://doi.org/10.1103/PhysRevLett.134.160404>
- [2] Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
- [3] Hiriart-Urruty, J.-B., Lemaréchal, C.: *Fundamentals of Convex Analysis*. Springer, Berlin (2001). <https://doi.org/10.1007/978-3-642-56468-0>
- [4] Shannon, C.E.: A mathematical theory of communication. *Bell System Technical Journal* **27**(3), 379–423 (1948) <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [5] Faddeev, D.K.: On the concept of entropy of a finite probabilistic scheme. *Uspekhi Matematicheskikh Nauk* **11**(1), 227–231 (1956)
- [6] Baez, J.C., Fritz, T., Leinster, T.: A characterization of entropy in terms of information loss. *Entropy* **13**, 1945–1957 (2011) <https://doi.org/10.3390/e13111945>
- [7] Jaynes, E.T.: Information theory and statistical mechanics. *Physical Review* **106**, 620–630 (1957) <https://doi.org/10.1103/PhysRev.106.620>
- [8] Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., Rustichini, A.: Axiomatic tests for the Boltzmann distribution. *Journal of Statistical Mechanics: Theory and Experiment* **2021**(1), 013406 (2021) <https://doi.org/10.1088/1742-5468/abd4ce>
- [9] Sandomirskiy, F., Tamuz, O.: Decomposable stochastic choice. In: *Proceedings of the 25th ACM Conference on Economics and Computation. EC '24*, p. 1195. Association for Computing Machinery, New York, NY, USA (2024). <https://doi.org/10.1145/3670865.3673590> . Full version: arXiv:2312.04827v2, 19 May 2024; Theorems 1 and 3. <https://arxiv.org/abs/2312.04827v2>
- [10] Sandomirskiy, F., Tamuz, O.: On the origin of the Boltzmann distribution. *Mathematische Annalen* **392**(4), 5617–5638 (2025) <https://doi.org/10.1007/s00208-025-03263-x>
- [11] Fu, W., Wang, Y., Zhang, Y., Zhao, H.: Boltzmann distribution from invariance of coarse-graining-scale and energy-shift. *Communications in Theoretical Physics* **78**, 115601 (2026) <https://doi.org/10.1088/1572-9494/ae8f21>
- [12] Hyers, D.H.: On the stability of the linear functional equation. *Proceedings of the National Academy of Sciences of the United States of America* **27**(4), 222–224 (1941) <https://doi.org/10.1073/pnas.27.4.222>
- [13] Boros, Z., Páles, Z., Volkman, P.: On stability for the Jensen equation on intervals. *Aequationes Mathematicae* **60**, 291–297 (2000) <https://doi.org/10.1007/s000100050155>
- [14] Hoeffding, W.: Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**(301), 13–30 (1963) <https://doi.org/10.1080/01621459.1963.10500830>
- [15] Campisi, M., Talkner, P., Hänggi, P.: Finite bath fluctuation theorem. *Physical Review E* **80**, 031145 (2009) <https://doi.org/10.1103/PhysRevE.80.031145>
- [16] Farías, C., Davis, S.: Temperature distribution in finite systems: Application to the one-dimensional Ising chain. *The European Physical Journal B* **96**, 39 (2023) <https://doi.org/10.1140/epjb/s10051-023-00509-y>
- [17] Jarzynski, C.: Nonequilibrium work theorem for a system strongly coupled to a thermal environment. *Journal of Statistical Mechanics: Theory and Experiment* **2004** (2004) <https://doi.org/10.1088/1742-5468/2004/09/P09005> . Article P09005
- [18] Talkner, P., Hänggi, P.: Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical. *Reviews of Modern Physics* **92**, 041002 (2020) <https://doi.org/10.1103/RevModPhys.92.041002>
- [19] Beck, C., Cohen, E.G.D.: Superstatistics. *Physica A* **322**, 267–275 (2003) [https://doi.org/10.1016/S0378-4371\(03\)00019-0](https://doi.org/10.1016/S0378-4371(03)00019-0)
- [20] Pusz, W., Woronowicz, S.L.: Passive states and KMS states for general quantum systems. *Communications in Mathematical Physics* **58**, 273–290 (1978) <https://doi.org/10.1007/BF01614224>