

A Rigorous Disorder-Adaptive External-Field Obstruction for Uniform Rapid Mixing in the SK Model

GPT-5.6 Sol

Abstract

We prove that the strongest simultaneous-over-all-external-fields rapid-mixing statement for the Sherrington–Kirkpatrick model is false below the usual high-temperature threshold. More precisely, at inverse temperature $\beta = 19/20$, with probability tending to one over the GOE disorder, there exists a finite-for-each- n , disorder-dependent, inhomogeneous external-field vector for which the single-site heat-bath Glauber dynamics has mixing time $\exp(\Omega(n))$.

The construction uses a fixed-depth Lanczos transcript. Conditional on that transcript, the GOE matrix decomposes exactly into a finite-rank path-like signal plus an independent GOE innovation projected onto the unexplored subspace. We retain the coordinates on which the top signal vector is large. The barrier side is controlled by an annealed Gaussian upper bound and a finite-dimensional entropy estimate. The two wells are certified by a transcript-adapted product trial measure and the Gibbs variational principle. This replaces a previously heuristic constrained quenched-free-energy expansion; no such expansion is used here.

1 Model and main theorem

Let W_n be an $n \times n$ GOE matrix with

$$(W_{ij})_{i < j} \stackrel{\text{i.i.d.}}{\sim} N(0, n^{-1}), \quad (W_{ii})_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} N(0, 2n^{-1}),$$

with the diagonal and upper-triangular entries mutually independent. For $\beta > 0$ and $h \in \mathbb{R}^n$, define

$$\mu_{W,h}(x) = \frac{1}{Z_{W,h}} \exp \left\{ \frac{\beta}{2} x^\top W x + h^\top x \right\}, \quad x \in \{\pm 1\}^n.$$

The diagonal of W contributes the constant $\frac{\beta}{2} \text{Tr } W$ and therefore does not affect the measure or the dynamics.

The discrete-time random-scan heat-bath chain chooses $i \in [n]$ uniformly and resamples X_i from its conditional law under $\mu_{W,h}$. Its total-variation mixing time is denoted by

$$t_{\text{mix}}(\varepsilon; W, h).$$

Theorem 1.1 (Disorder-adaptive field producing exponential slow mixing). *Set*

$$\beta = \frac{19}{20}.$$

There exists a constant $c > 0$ such that

$$\mathbb{P}_{W_n} \left(\exists h \in \mathbb{R}^n : t_{\text{mix}} \left(\frac{1}{4}; W_n, h \right) \geq e^{cn} \right) \longrightarrow 1.$$

Consequently, for every fixed $K < \infty$,

$$\mathbb{P}_{W_n} \left(\forall h \in \mathbb{R}^n : t_{\text{mix}} \left(\frac{1}{4}; W_n, h \right) \leq n^K \right) \longrightarrow 0.$$

Overview. The basic mechanism is spectral: a sufficiently strong positive mode creates two free-energy wells at positive and negative overlap, while configurations near zero overlap lose an order- n amount of free energy. Once this gap is linear in n , the separating region has mass $e^{-\Theta(n)}$, giving exponential slow mixing for single-site dynamics.

We realize this mechanism by restricting to an active set where a leading mode remains strong, proving a two-well/barrier structure there, and then pinning the complementary spins so that the active model is recovered exactly as a conditional subsystem. A fixed-depth Lanczos exploration and Gaussian regression isolate a path-like finite-rank signal plus an independent projected GOE remainder; Haar truncation selects the active coordinates, while product-measure lower bounds and an annealed entropy bound separate the wells from the barrier. A strong disorder-dependent external field then embeds this bottleneck into the full chain, yielding exponentially small conductance.

Remark 1.2 (Quantifier distinction). *The field constructed below is an inhomogeneous vector that depends on the realized disorder, and its largest coordinates grow with n . The theorem therefore disproves the simultaneous statement*

$$\mathbb{P}_W(\text{the desired estimate holds for every } h) \rightarrow 1.$$

It does not disprove the weaker statement in which a deterministic field sequence h_n is fixed before W_n is sampled, nor a statement restricted to uniformly bounded or spatially homogeneous fields.

The formal proof is divided into four parts:

- (i) an exact finite-query Gaussian decomposition of W ;
- (ii) a disorder-adaptive active set carrying a finite-rank signal;
- (iii) an exponentially small free-energy barrier for the zero-field active subsystem;
- (iv) embedding that subsystem into the original model by a finite-for-each- n external field.

2 Finite-depth Lanczos decomposition

Until the final projection argument in the proof of Theorem 1.1, all probability statements are with respect to the joint law of (W, q_0) .

Fix an integer $L \geq 1$ and assume $n > L + 1$. Let q_0 be Haar-uniform on the unit sphere and independent of W . Define the Lanczos recursion

$$q_{-1} = 0, \quad b_0 = 0,$$

and, for $k = 0, \dots, L - 1$,

$$\begin{aligned} a_k &= q_k^\top W q_k, \\ r_k &= W q_k - a_k q_k - b_k q_{k-1}, \\ b_{k+1} &= \|r_k\|_2, \\ q_{k+1} &= r_k / b_{k+1}. \end{aligned}$$

The Lanczos vectors are orthonormal. At step k , conditional on the preceding transcript, r_k is a nondegenerate Gaussian vector in an $(n - k - 1)$ -dimensional space; hence $b_{k+1} > 0$ almost surely and breakdown has probability zero. For $0 \leq k \leq L$, put

$$Q_k = (q_0, \dots, q_{k-1}), \quad P_k = Q_k Q_k^\top,$$

with $P_0 = 0$. The full depth- L transcript is

$$\mathcal{T}_L = (q_0, \dots, q_L; a_0, \dots, a_{L-1}; b_1, \dots, b_L).$$

Lemma 2.1 (Exact adaptive Gaussian regression). *Conditional on $\mathcal{T}_L$, the conditional law of W admits the representation*

$$W \stackrel{d}{=} M_L + P_L^\perp G P_L^\perp,$$

where G is a standard GOE matrix independent of $\mathcal{T}_L$, and

$$\begin{aligned} M_L = & \sum_{k=0}^{L-1} a_k q_k q_k^\top + \sum_{k=1}^{L-1} b_k (q_{k-1} q_k^\top + q_k q_{k-1}^\top) \\ & + b_L (q_{L-1} q_L^\top + q_L q_{L-1}^\top). \end{aligned} \quad (1)$$

Moreover, conditional on the tridiagonal coefficients $(a_0, \dots, a_{L-1}, b_1, \dots, b_L)$, the frame $(q_0, \dots, q_L)$ is Haar on the Stiefel manifold. In the preceding representation it is independent of G .

Proof. We first record the fixed-subspace conditioning formula. Let $Q \in \mathbb{R}^{n \times r}$ have orthonormal columns, let $P_Q = Q Q^\top$, and suppose that $Y = W Q$ is observed. Complete Q to an orthogonal matrix $(Q, Q_\perp)$. In this basis,

$$(Q, Q_\perp)^\top W (Q, Q_\perp) = \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix}, \quad Y = Q A + Q_\perp B.$$

By orthogonal invariance of the GOE, the Gaussian blocks A , B , and C are independent except for the symmetry within A and C . Thus conditioning on Y fixes A and B and leaves C as an independent GOE block with the original $1/n$ scaling. Rewriting the fixed blocks in the original coordinates gives

$$W \mid \{W Q = Y\} \stackrel{d}{=} Y Q^\top + Q Y^\top - Q (Q^\top Y) Q^\top + P_Q^\perp G P_Q^\perp, \quad (2)$$

where G is an independent standard GOE matrix. The first three terms are the unique symmetric matrix with the prescribed Q -columns and with zero $Q^\perp$ - $Q^\perp$ block.

We now explain why the same formula may be iterated for adaptive Lanczos queries. Let

$$\mathcal{F}_k = \sigma(q_0, \dots, q_k; a_0, \dots, a_{k-1}; b_1, \dots, b_k), \quad 0 \leq k \leq L.$$

We prove inductively that, conditional on $\mathcal{F}_k$,

$$W \stackrel{d}{=} M_k + P_k^\perp G_k P_k^\perp,$$

where G_k is a fresh GOE matrix independent of $\mathcal{F}_k$, and M_k reproduces the already revealed vectors $W q_0, \dots, W q_{k-1}$. The case $k = 0$ follows because q_0 is independent of W . For the induction step, q_k is $\mathcal{F}_k$ -measurable and lies in $\text{range}(P_k^\perp)$. The next transcript increment reveals the full vector

$$W q_k = b_k q_{k-1} + a_k q_k + b_{k+1} q_{k+1}.$$

After subtracting the known contribution $M_k q_k$, this is one fixed-vector query of the remaining Gaussian matrix $P_k^\perp G_k P_k^\perp$. Applying (2) inside $\text{range}(P_k^\perp)$ leaves a fresh independent GOE matrix on the orthogonal complement of $\text{span}(q_0, \dots, q_k)$, namely $P_{k+1}^\perp G_{k+1} P_{k+1}^\perp$. This proves the induction. At $k = L$, collecting the revealed actions gives exactly the matrix M_L in (1).

Finally, the joint law of (W, q_0) is invariant under

$$(W, q_0) \mapsto (UWU^\top, Uq_0), \quad U \in O(n).$$

The coefficients (a_k, b_k) are unchanged, whereas the frame is sent to $(Uq_0, \dots, Uq_L)$. Hence, conditional on the coefficients, the frame has a left-orthogonally invariant law on the Stiefel manifold. By uniqueness of Haar probability measure, this law is Haar. The independence from G follows from the Gaussian innovations used in the preceding induction. $\square$

Lemma 2.2 (Fixed-depth path limit). *For every fixed L ,*

$$\max_{0 \leq k \leq L-1} |a_k| \xrightarrow{\mathbb{P}} 0, \quad \max_{1 \leq k \leq L} |b_k - 1| \xrightarrow{\mathbb{P}} 0.$$

Consequently, the action of M_L on

$$\mathcal{Q}_{L+1} := \text{span}(q_0, \dots, q_L)$$

converges in operator norm to the adjacency matrix of the path on $L+1$ vertices. If its eigenvalues are ordered as

$$\rho_{1,n} \geq \rho_{2,n} \geq \dots \geq \rho_{L+1,n},$$

then

$$\rho_{k,n} \xrightarrow{\mathbb{P}} \rho_k^{(L)} := 2 \cos \frac{k\pi}{L+2}.$$

Proof. The induction in Lemma 2.1 also gives the standard finite-dimensional GOE tridiagonal law:

$$a_k \sim N(0, 2/n), \quad b_{k+1} \stackrel{d}{=} \frac{1}{\sqrt{n}} \chi_{n-k-1},$$

and these newly exposed coefficients are mutually independent. Indeed, the component of the fresh Gaussian query along q_k has variance $2/n$, while its component orthogonal to $q_0, \dots, q_k$ consists of $n-k-1$ independent $N(0, 1/n)$ coordinates. Since L is fixed,

$$a_k = O_{\mathbb{P}}(n^{-1/2}), \quad b_{k+1} = 1 + O_{\mathbb{P}}(n^{-1/2})$$

for every $k \leq L-1$, and the maxima over finitely many k have the same limits.

In the ordered basis $(q_0, \dots, q_L)$, the restriction of M_L is the tridiagonal matrix with diagonal $(a_0, \dots, a_{L-1}, 0)$ and off-diagonal $(b_1, \dots, b_L)$. It therefore converges in operator norm to the path adjacency matrix with zero diagonal and unit off-diagonal. The latter has normalized sine eigenvectors and eigenvalues $2 \cos(k\pi/(L+2))$, $1 \leq k \leq L+1$. Weyl's inequality gives the claimed eigenvalue convergence. $\square$

Let $z_{1,n}, \dots, z_{L+1,n}$ be an orthonormal eigenbasis of M_L in $\mathcal{Q}_{L+1}$, with $z_{1,n}$ corresponding to $\rho_{1,n}$. Since every $b_k > 0$, the tridiagonal restriction is an irreducible Jacobi matrix and has simple spectrum. We choose the eigenvector signs by any fixed measurable convention. Conditional on the coefficients, this eigenbasis is a deterministic orthogonal rotation of the Haar frame $(q_0, \dots, q_L)$. In particular, $z_{1,n}$ is Haar, and the full eigenframe is independent of the GOE innovation G in Lemma 2.1.

3 Haar truncation and the active subsystem

Fix a threshold $a > 0$ and set

$$S_n = S_n(a) := \{i : \sqrt{n} |z_{1,n}(i)| > a\}.$$

Let P_{S_n} denote coordinate projection onto S_n . Define

$$p(a) := \mathbb{P}(|Z| > a), \quad s(a) := \mathbb{E}[Z^2 \mathbf{1}_{\{|Z| > a\}}], \quad \ell(a) := \mathbb{E}[|Z| \mathbf{1}_{\{|Z| > a\}}] = 2\varphi(a),$$

where $Z \sim N(0, 1)$ and φ is the standard Gaussian density.

Lemma 3.1 (Truncated Haar-frame laws). *For fixed L and a ,*

$$\frac{|S_n|}{n} \xrightarrow{\mathbb{P}} p(a), \tag{3}$$

$$z_{1,n}^\top P_{S_n} z_{1,n} \xrightarrow{\mathbb{P}} s(a), \tag{4}$$

$$\frac{1}{\sqrt{n}} \sum_{i \in S_n} |z_{1,n}(i)| \xrightarrow{\mathbb{P}} \ell(a), \tag{5}$$

$$\max_i |z_{1,n}(i)| = O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{n}} \right). \tag{6}$$

Moreover, for $k, l \geq 2$,

$$z_{k,n}^\top P_{S_n} z_{l,n} \xrightarrow{\mathbb{P}} p(a) \mathbf{1}_{\{k=l\}}, \tag{7}$$

$$z_{1,n}^\top P_{S_n} z_{k,n} \xrightarrow{\mathbb{P}} 0, \tag{8}$$

and, with $b_i = \text{sgn}(z_{1,n}(i)) \mathbf{1}_{\{i \in S_n\}}$,

$$\frac{1}{\sqrt{n}} z_{k,n}^\top b \xrightarrow{\mathbb{P}} 0, \quad k \geq 2. \tag{9}$$

Proof. Conditional on the Lanczos coefficients, $z_{1,n}$ is Haar on the unit sphere. Hence we may write

$$z_{1,n} = \frac{g}{\|g\|_2}, \quad g_1, \dots, g_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1),$$

and put $\alpha_n = \|g\|_2 / \sqrt{n}$. Then $\alpha_n \rightarrow 1$ in probability and

$$S_n = \{i : |g_i| > a\alpha_n\}.$$

The random threshold causes no difficulty. For every fixed $\gamma > 0$, on the event $|\alpha_n - 1| \leq \gamma$,

$$\mathbf{1}_{\{|g_i| > a(1+\gamma)\}} \leq \mathbf{1}_{\{i \in S_n\}} \leq \mathbf{1}_{\{|g_i| > a(1-\gamma)\}}.$$

The law of large numbers applied to the two deterministic thresholds, followed by $\gamma \downarrow 0$, proves (3); the continuity of the Gaussian distribution at $\pm a$ is the required anti-concentration input. The same sandwich argument, now with the integrable weights g_i^2 and $|g_i|$, gives

$$\frac{n^{-1} \sum_i g_i^2 \mathbf{1}_{\{i \in S_n\}}}{n^{-1} \sum_i g_i^2} \xrightarrow{\mathbb{P}} s(a)$$

and

$$\frac{n^{-1} \sum_i |g_i| \mathbf{1}_{\{i \in S_n\}}}{\alpha_n} \xrightarrow{\mathbb{P}} \ell(a),$$

which are exactly (4) and (5). Finally, a Gaussian union bound gives $\max_i |g_i| = O_{\mathbb{P}}(\sqrt{\log n})$; since $\alpha_n \rightarrow 1$, this proves (6).

We next control the remaining vectors in the Haar frame. Condition on $z_{1,n}$ and write

$$\Pi = I - z_{1,n} z_{1,n}^\top, \quad H = \Pi P_{S_n} \Pi.$$

The vectors $z_{2,n}, \dots, z_{L+1,n}$ form a fixed-dimensional Haar frame in the $(n-1)$ -dimensional space $z_{1,n}^\perp$. Standard spherical second-moment identities give, for each fixed $k \geq 2$,

$$\mathbb{E}[z_{k,n}^\top H z_{k,n} \mid z_{1,n}] = \frac{\text{Tr } H}{n-1}, \quad \text{Var}(z_{k,n}^\top H z_{k,n} \mid z_{1,n}) \leq \frac{C \text{Tr}(H^2)}{(n-1)^2},$$

and, for $k \neq l$,

$$\mathbb{E}[(z_{k,n}^\top H z_{l,n})^2 \mid z_{1,n}] \leq \frac{C \text{Tr}(H^2)}{(n-1)^2}.$$

Here $\text{Tr}(H^2) \leq \text{Tr } H \leq n$, while

$$\frac{\text{Tr } H}{n-1} = \frac{|S_n| - z_{1,n}^\top P_{S_n} z_{1,n}}{n-1} \xrightarrow{\mathbb{P}} p(a).$$

Chebyshev's inequality therefore proves (7) for the finite collection of indices under consideration.

For the mixed term, conditional spherical symmetry gives

$$\mathbb{E}[(z_{1,n}^\top P_{S_n} z_{k,n})^2 \mid z_{1,n}] = \frac{\|\Pi P_{S_n} z_{1,n}\|_2^2}{n-1} \leq \frac{1}{n-1},$$

which proves (8). Finally, b is measurable with respect to $z_{1,n}$ and $\|b\|_2^2 = |S_n| \leq n$, so

$$\mathbb{E} \left[\left(\frac{z_{k,n}^\top b}{\sqrt{n}} \right)^2 \mid z_{1,n} \right] = \frac{\|\Pi b\|_2^2}{n(n-1)} \leq \frac{1}{n-1}.$$

This proves (9). □

We now fix

$$L = 20, \quad a = \frac{4}{5}.$$

Conditional on the Lanczos transcript, the zero-field active Hamiltonian on $\{\pm 1\}^{S_n}$ is

$$\mathcal{H}_n(x) = \frac{\beta}{2} x^\top W_{S_n, S_n} x = \frac{\beta}{2} x^\top M_{L, S_n} x + \frac{\beta}{2} x^\top R_{S_n} x, \quad (10)$$

where

$$R = P_L^\perp G P_L^\perp.$$

Here G is independent of the transcript; the projected matrix R itself depends on the transcript through P_L . Conditional on the transcript, R is the centered Gaussian remainder on the unexplored subspace. Diagonal terms are configuration-independent constants and are harmless on the exponential scale used below.

For $1 \leq k \leq L+1$, define the finite-dimensional order parameters

$$q_k(x) := \frac{z_{k,n}^\top x}{\sqrt{n}}, \quad x \in \{\pm 1\}^{S_n},$$

where x is extended by zero outside S_n . Since

$$M_L = \sum_{k=1}^{L+1} \rho_{k,n} z_{k,n} z_{k,n}^\top$$

on $\mathcal{Q}_{L+1}$, we obtain

$$\frac{1}{n} x^\top M_{L,S_n} x = \sum_{k=1}^{L+1} \rho_{k,n} q_k(x)^2. \quad (11)$$

Thus q_1 measures the overlap with the strongest retained signal direction, while $q_2, \dots, q_{L+1}$ record the remaining finite-rank directions.

4 Barrier upper bound

We first prove a finite-dimensional entropy estimate. Its role is to show that, once q_1 is constrained near zero, the entropy cannot exploit the remaining finite-rank directions enough to create another macroscopic gain.

Lemma 4.1 (Quadratic entropy upper bound). *Fix $d \geq 2$. For each n , let $u_{1,n}, \dots, u_{d,n} \in \mathbb{R}^n$ be orthonormal and let $S_n \subseteq [n]$. Define*

$$A_n = (u_{k,n}^\top P_{S_n} u_{l,n})_{k,l \leq d}$$

and, for $x \in \{\pm 1\}^{S_n}$,

$$q(x) = \frac{1}{\sqrt{n}} (u_{1,n}^\top x, \dots, u_{d,n}^\top x).$$

Assume $A_n \rightarrow A \succ 0$, and let $D_n \rightarrow D$ be diagonal matrices. If the lower-right $(d-1) \times (d-1)$ block of

$$B := A^{-1} - \beta D$$

is positive definite, then there exist constants $\varepsilon_0 > 0$ and $C < \infty$ such that, for every fixed $0 < \varepsilon \leq \varepsilon_0$,

$$\limsup_{n \rightarrow \infty} \left\{ \frac{1}{n} \log \sum_{\substack{x \in \{\pm 1\}^{S_n} \\ |q_1(x)| \leq \varepsilon}} \exp \left\{ \frac{\beta n}{2} q(x)^\top D_n q(x) \right\} - \frac{|S_n|}{n} \log 2 \right\} \leq C \varepsilon^2. \quad (12)$$

Proof. For $t \in \mathbb{R}^d$,

$$\begin{aligned} \sum_{x \in \{\pm 1\}^{S_n}} e^{nt^\top q(x)} &= 2^{|S_n|} \prod_{i \in S_n} \cosh \left(\sqrt{n} \sum_{k=1}^d t_k u_{k,n}(i) \right) \\ &\leq 2^{|S_n|} \exp \left\{ \frac{n}{2} t^\top A_n t \right\}, \end{aligned} \quad (13)$$

where we used $\log \cosh y \leq y^2/2$. Also $|q_k(x)| \leq 1$, so the range of $q(x)$ is contained in the fixed cube $[-1, 1]^d$.

Partition this cube into cubes of side $\delta_n = n^{-2}$. Their number is at most $(2n^2 + 2)^d$, whose logarithm is $o(n)$. For a cube centered at $r \in [-1, 1]^d$, let $N_n(r)$ be the number of configurations whose $q(x)$ lies in that cube. Since $A_n \rightarrow A \succ 0$, the matrices A_n^{-1} are uniformly bounded for all sufficiently large n . Set $t = A_n^{-1}r$. For every x in the cube,

$$t^\top q(x) \geq t^\top r - \|t\|_1 \delta_n.$$

Combining this with (13) gives, uniformly over all cubes,

$$\begin{aligned} \frac{1}{n} \log N_n(r) &\leq \frac{|S_n|}{n} \log 2 - t^\top r + \frac{1}{2} t^\top A_n t + O(\delta_n) \\ &= \frac{|S_n|}{n} \log 2 - \frac{1}{2} r^\top A_n^{-1} r + o(1). \end{aligned}$$

The uniformity follows because both r and $A_n^{-1}r$ range over bounded sets.

The quadratic energy also varies uniformly by $o(1)$ on the $1/n$ logarithmic scale within each cube: since D_n is uniformly bounded,

$$\sup_{q, r \text{ in the same cube}} \left| \frac{\beta}{2} q^\top D_n q - \frac{\beta}{2} r^\top D_n r \right| = O(\delta_n).$$

Summing over the polynomial number of cubes therefore yields

$$\begin{aligned} &\frac{1}{n} \log \sum_{\substack{x \in \{\pm 1\}^{S_n} \\ |q_1(x)| \leq \varepsilon}} \exp \left\{ \frac{\beta n}{2} q(x)^\top D_n q(x) \right\} \\ &\leq \frac{|S_n|}{n} \log 2 + \sup_{|r_1| \leq \varepsilon + o(1)} \left\{ -\frac{1}{2} r^\top (A_n^{-1} - \beta D_n) r \right\} + o(1). \end{aligned} \tag{14}$$

It remains to optimize the finite-dimensional quadratic form. Write

$$B_n := A_n^{-1} - \beta D_n = \begin{pmatrix} \alpha_n & v_n^\top \\ v_n & C_n \end{pmatrix}, \quad r = \begin{pmatrix} r_1 \\ y \end{pmatrix}.$$

By assumption, the limiting lower-right block is positive definite, so for some $\kappa > 0$ and all sufficiently large n , $C_n \succeq \kappa I$. Completing the square gives

$$\begin{aligned} -\frac{1}{2} r^\top B_n r &= -\frac{1}{2} (y + C_n^{-1} v_n r_1)^\top C_n (y + C_n^{-1} v_n r_1) \\ &\quad + \frac{1}{2} (v_n^\top C_n^{-1} v_n - \alpha_n) r_1^2 \leq C r_1^2, \end{aligned}$$

where C is uniform because B_n and C_n^{-1} are uniformly bounded. Substitution into (14) proves (12). The same estimates are uniform when A_n and D_n remain in a sufficiently small fixed neighborhood of A and D ; this uniform form will be used below on a high-probability transcript event. $\square$

Proposition 4.2 (Annealed upper bound on the separating slab). *Let*

$$\mathcal{B}_n(\varepsilon) := \{x \in \{\pm 1\}^{S_n} : |q_1(x)| \leq \varepsilon\}$$

and

$$Z_{\mathcal{B}_n} := \sum_{x \in \mathcal{B}_n(\varepsilon)} e^{\mathcal{H}_n(x)}.$$

There exist constants $\varepsilon_0 > 0$ and $C < \infty$ such that, for every fixed $0 < \varepsilon \leq \varepsilon_0$ and every $\eta > 0$, with probability tending to one,

$$\frac{1}{n} \log Z_{\mathcal{B}_n} \leq p(a) \log 2 + \frac{\beta^2 p(a)^2}{4} + C\varepsilon^2 + \eta. \quad (15)$$

Proof. Condition on the Lanczos transcript. For fixed $x \in \{\pm 1\}^{S_n}$, extended by zero outside S_n , let

$$u = P_L^\perp P_{S_n} x.$$

Then $x^\top R_{S_n} x = u^\top G u$. The GOE covariance identity gives

$$\text{Var}(u^\top G u) = \frac{2}{n} \|u\|_2^4.$$

Since $u^\top G u$ is centered Gaussian,

$$\mathbb{E}_G \exp \left\{ \frac{\beta}{2} u^\top G u \right\} = \exp \left\{ \frac{\beta^2}{4n} \|u\|_2^4 \right\} \leq \exp \left\{ \frac{\beta^2 |S_n|^2}{4n} \right\},$$

where $\|u\|_2^2 \leq \|P_{S_n} x\|_2^2 = |S_n|$. Consequently,

$$\mathbb{E}_G Z_{\mathcal{B}_n} \leq \exp \left\{ \frac{\beta^2 |S_n|^2}{4n} \right\} \sum_{x \in \mathcal{B}_n(\varepsilon)} \exp \left\{ \frac{\beta}{2} x^\top M_{L, S_n} x \right\}. \quad (16)$$

Apply Lemma 4.1 with $d = L + 1$, $u_{k,n} = z_{k,n}$, and

$$D_n = \text{diag}(\rho_{1,n}, \dots, \rho_{L+1,n}).$$

By Lemmas 2.2 and 3.1,

$$A_n \xrightarrow{\mathbb{P}} A := \text{diag}(s(a), p(a), \dots, p(a)), \quad D_n \xrightarrow{\mathbb{P}} D := \text{diag}(\rho_1^{(L)}, \dots, \rho_{L+1}^{(L)}).$$

The limiting lower-right block of $A^{-1} - \beta D$ is diagonal, with entries

$$\frac{1}{p(a)} - \beta \rho_k^{(L)}, \quad k \geq 2.$$

Because $\rho_k^{(L)} \leq \rho_2^{(L)}$ for $k \geq 2$, it is positive definite precisely when

$$\frac{1}{p(a)} - \beta \rho_2^{(L)} > 0.$$

Lemma 6.1 verifies this inequality with a strict margin. Hence there is a transcript-measurable good event of probability tending to one on which the lower-right block remains uniformly positive definite and the constant C in Lemma 4.1 may be chosen uniformly. On that event, (16) gives

$$\frac{1}{n} \log \mathbb{E}_G Z_{\mathcal{B}_n} \leq \frac{|S_n|}{n} \log 2 + \frac{\beta^2}{4} \left(\frac{|S_n|}{n} \right)^2 + C\varepsilon^2 + o(1).$$

Using $|S_n|/n \rightarrow p(a)$, for every $\eta > 0$ this is at most

$$p(a) \log 2 + \frac{\beta^2 p(a)^2}{4} + C\varepsilon^2 + \frac{\eta}{2}$$

for all sufficiently large n on another event of probability tending to one. Conditional Markov's inequality then gives

$$\mathbb{P}_G \left(\frac{1}{n} \log Z_{\mathcal{B}_n} > p(a) \log 2 + \frac{\beta^2 p(a)^2}{4} + C\varepsilon^2 + \eta \mid \mathcal{T}_L \right) \leq e^{-\eta n/2}.$$

Averaging over the transcript proves (15). □

5 A product-measure lower bound for the wells

Fix

$$\delta = \frac{17}{20}.$$

Let

$$b_i = \text{sgn}(z_{1,n}(i)), \quad i \in S_n,$$

and define the product measure ν_+ on $\{\pm 1\}^{S_n}$ by

$$\nu_+(X_i = 1) = \frac{1 + \delta b_i}{2}.$$

Conditional on the Lanczos transcript, this is a deterministic product measure and is independent of the GOE innovation G . Its mean vector is $m = \delta b$. Define

$$h_2 \left(\frac{1 + \delta}{2} \right) = \log 2 - I(\delta),$$

where

$$I(\delta) = \frac{1}{2} [(1 + \delta) \log(1 + \delta) + (1 - \delta) \log(1 - \delta)]. \quad (17)$$

Proposition 5.1 (Variational lower bound on each well). *Set*

$$q_* := \frac{1}{4} \delta \ell(a) > 0.$$

For every $\eta > 0$, with probability tending to one,

$$\frac{1}{n} \log \sum_{\substack{x \in \{\pm 1\}^{S_n} \\ q_1(x) \geq q_*}} e^{\mathcal{H}_n(x)} \geq p(a) \log 2 - p(a) I(\delta) + \frac{\beta \rho_1^{(L)}}{2} \delta^2 \ell(a)^2 - \eta. \quad (18)$$

The same estimate holds with $q_1(x) \leq -q_$.*

Proof. Under ν_+ ,

$$\mathbb{E}_{\nu_+} q_1(X) = \frac{\delta}{\sqrt{n}} \sum_{i \in S_n} |z_{1,n}(i)| \xrightarrow{\mathbb{P}} \delta \ell(a).$$

We make the concentration step explicit. Put $w_i = z_{1,n}(i)/\sqrt{n}$. The independent summands $w_i X_i$ have ranges of length $2|w_i|$, and

$$\sum_{i \in S_n} w_i^2 = \frac{1}{n} z_{1,n}^\top P_{S_n} z_{1,n} = \frac{s(a) + o_{\mathbb{P}}(1)}{n}.$$

On a transcript event of probability tending to one, the mean is at least $3\delta\ell(a)/4$ and $z_{1,n}^\top P_{S_n} z_{1,n}$ is bounded by a fixed constant. Hoeffding's inequality therefore gives

$$\begin{aligned} \nu_+ \left(q_1(X) \leq \frac{1}{2} \delta\ell(a) \right) &\leq \exp \left\{ -\frac{n(\delta\ell(a)/4)^2}{2z_{1,n}^\top P_{S_n} z_{1,n}} \right\} \\ &\leq e^{-c_0 n} \end{aligned}$$

for some deterministic $c_0 > 0$. Since $q_* < \delta\ell(a)/2$, the event

$$E_n := \{q_1 \geq q_*\}$$

has ν_+ -probability $1 - \alpha_n$ with $\alpha_n \leq e^{-c_0 n}$.

Let $\tilde{\nu}_+ = \nu_+(\cdot \mid E_n)$. The entropy loss caused by this conditioning is negligible. Indeed, the entropy decomposition across E_n and E_n^c gives

$$H(\nu_+) = h_2(\alpha_n) + (1 - \alpha_n)H(\tilde{\nu}_+) + \alpha_n H(\nu_+(\cdot \mid E_n^c)),$$

so, because every entropy on $\{\pm 1\}^{S_n}$ is at most $|S_n| \log 2$,

$$H(\tilde{\nu}_+) = H(\nu_+) + o(n).$$

Moreover, $\|\tilde{\nu}_+ - \nu_+\|_{\text{TV}} = \alpha_n$. On the event $\|W\|_{\text{op}} = O(1)$,

$$\sup_{x \in \{\pm 1\}^{S_n}} |\mathcal{H}_n(x)| \leq \frac{\beta}{2} \|W\|_{\text{op}} |S_n| = O(n),$$

and hence

$$\mathbb{E}_{\tilde{\nu}_+} \mathcal{H}_n = \mathbb{E}_{\nu_+} \mathcal{H}_n + o(n).$$

The Gibbs variational principle on the restricted state space E_n therefore yields

$$\log Z_{\{q_1 \geq q_*\}} \geq H(\tilde{\nu}_+) + \mathbb{E}_{\tilde{\nu}_+} \mathcal{H}_n = H(\nu_+) + \mathbb{E}_{\nu_+} \mathcal{H}_n - o(n). \quad (19)$$

The entropy is exactly

$$H(\nu_+) = |S_n|(\log 2 - I(\delta)).$$

For the finite-rank part, independence of the coordinates under ν_+ gives

$$\mathbb{E}_{\nu_+} [X^\top M_{L,S_n} X] = m^\top M_{L,S_n} m + (1 - \delta^2) \text{Tr}(P_{S_n} M_L P_{S_n}).$$

The trace term is $O(1)$ because $P_{S_n} M_L P_{S_n}$ has rank at most $L + 1$ and bounded operator norm. Also,

$$\frac{1}{n} m^\top M_{L,S_n} m = \delta^2 \sum_{k=1}^{L+1} \rho_{k,n} \left(\frac{z_{k,n}^\top b}{\sqrt{n}} \right)^2.$$

For $k = 1$,

$$\frac{z_{1,n}^\top b}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i \in S_n} |z_{1,n}(i)| \xrightarrow{\mathbb{P}} \ell(a),$$

whereas (9) makes every term with $k \geq 2$ vanish. Therefore

$$\frac{1}{n} m^\top M_{L,S_n} m \xrightarrow{\mathbb{P}} \rho_1^{(L)} \delta^2 \ell(a)^2.$$

For the residual part,

$$\mathbb{E}_{\nu_+}[X^\top R_{S_n} X] = m^\top R m + (1 - \delta^2) \text{Tr}(P_{S_n} R P_{S_n}).$$

Conditional on the transcript, both terms are centered Gaussian linear or quadratic functionals of the independent matrix G . With $u = P_L^\perp m$,

$$\text{Var}(m^\top R m) = \frac{2}{n} \|u\|_2^4 \leq \frac{2}{n} \|m\|_2^4 = O(n).$$

Also, with $A = P_L^\perp P_{S_n} P_L^\perp$,

$$\text{Tr}(P_{S_n} R P_{S_n}) = \text{Tr}(GA)$$

and

$$\text{Var}(\text{Tr}(GA)) = \frac{2}{n} \text{Tr}(A^2) \leq \frac{2}{n} \text{Tr}(A) = O(1).$$

Thus the full residual expectation is $o_{\mathbb{P}}(n)$. Dividing (19) by n , using $|S_n|/n \rightarrow p(a)$, and collecting the preceding limits gives (18) for every fixed $\eta > 0$ with probability tending to one. The negative well has the same partition function by the exact spin-flip symmetry $x \mapsto -x$. $\square$

6 Strict numerical separation

Lemma 6.1 (Certified parameter inequalities). *For*

$$\beta = \frac{19}{20}, \quad L = 20, \quad a = \frac{4}{5}, \quad \delta = \frac{17}{20},$$

one has

$$\beta \rho_2^{(20)} p(a) < 0.773 < 1, \tag{20}$$

$$\frac{\beta \rho_1^{(20)}}{2} \delta^2 \ell(a)^2 - p(a) I(\delta) - \frac{\beta^2 p(a)^2}{4} > 0.0067. \tag{21}$$

Proof. The following strict interval bounds suffice:

$$\begin{aligned} 1.9795 &< 2 \cos(\pi/22) < 1.9797, \\ 1.9189 &< 2 \cos(\pi/11) < 1.9191, \\ 0.42370 &< p(4/5) < 0.42372, \\ 0.57938 &< \ell(4/5) < 0.57939, \\ 0.42676 &< I(17/20) < 0.42677. \end{aligned}$$

For completeness, these intervals may be checked using only rational arithmetic. Start from the rational enclosures

$$\begin{aligned} 3.141592653589793238462643383279 &< \pi < 3.141592653589793238462643383280, \\ 1.414213562373095048801688724209 &< \sqrt{2} < 1.414213562373095048801688724210, \\ 1.772453850905516027298167483341 &< \sqrt{\pi} < 1.772453850905516027298167483342. \end{aligned}$$

The first can, for example, be certified from Machin's formula; the latter two are then verified by squaring. Evaluate the alternating cosine series through the $x^{12}/12!$ term; for $0 < x \leq \pi/11$, the omitted term is below 3×10^{-19} . For

$$p(4/5) = 1 - \operatorname{erf}\left(\frac{4}{5\sqrt{2}}\right),$$

use

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{k \geq 0} \frac{(-1)^k x^{2k+1}}{k!(2k+1)}$$

through $k = 8$; the next term bounds the remainder by 3.3×10^{-12} . For

$$\ell(4/5) = \sqrt{\frac{2}{\pi}} e^{-8/25},$$

use the alternating exponential series through degree 10, whose remainder is below 9.1×10^{-14} . Finally, for

$$\log y = 2 \sum_{r \geq 0} \frac{1}{2r+1} \left(\frac{y-1}{y+1} \right)^{2r+1},$$

truncate after $r = 24$. The tail is bounded by

$$\frac{2|z|^{51}}{51(1-z^2)}, \quad z = \frac{y-1}{y+1},$$

which is below 1.8×10^{-8} for the slower case $y = 3/20$ and much smaller for $y = 37/20$. Outward-rounded rational interval arithmetic then gives the five displayed intervals.

Using the unfavorable endpoints,

$$\begin{aligned} &\beta(1.9191)(0.42372) < 0.773, \\ &\frac{0.95(1.9795)}{2}(0.85)^2(0.57938)^2 - (0.42372)(0.42677) - \frac{(0.95)^2(0.42372)^2}{4} > 0.0067. \end{aligned}$$

This proves both claims. □

7 An exponentially small bottleneck

Let Z_n be the full active partition function and let μ_n^{act} be the zero-field active Gibbs measure defined by (10).

Proposition 7.1 (Exponential barrier and active slow mixing). *There exist fixed constants $\varepsilon > 0$ and $c_1 > 0$ such that, with probability tending to one,*

$$\mu_n^{\text{act}}(|q_1| \leq \varepsilon) \leq e^{-c_1 n}.$$

Consequently, the random-scan heat-bath chain on the active subsystem satisfies

$$t_{\text{mix}}^{\text{act}}(1/4) \geq e^{c_1 n/2}$$

for all sufficiently large n on the same high-probability event.

Proof. Let

$$\Delta_* := \frac{\beta \rho_1^{(20)}}{2} \delta^2 \ell(a)^2 - p(a)I(\delta) - \frac{\beta^2 p(a)^2}{4} > 0.$$

By Lemma 6.1, $\Delta_* > 0.0067$. Let ε_0 and C be as in Proposition 4.2. Choose $\eta > 0$ and then a fixed

$$0 < \varepsilon < \min\{\varepsilon_0, q_*\}$$

so small that

$$C\varepsilon^2 + 2\eta < \Delta_*/2.$$

Apply Proposition 4.2 with error η and Proposition 5.1 with error η . Their exponent difference is at most

$$-\Delta_* + C\varepsilon^2 + 2\eta < -\Delta_*/2.$$

Hence, with probability tending to one,

$$\frac{Z_{\mathcal{B}_n(\varepsilon)}}{Z_{\{q_1 \geq q_*\}}} \leq e^{-c_1 n}$$

for some deterministic $c_1 > 0$. Since the full active partition function is at least $Z_{\{q_1 \geq q_*\}}$,

$$\mu_n^{\text{act}}(|q_1| \leq \varepsilon) \leq e^{-c_1 n}.$$

Set

$$A_n = \{x : q_1(x) < 0\}.$$

Spin-flip symmetry gives

$$\mu_n^{\text{act}}(A_n) = \frac{1}{2} (1 - \mu_n^{\text{act}}(q_1 = 0)) = \frac{1}{2} + o(1),$$

and in particular $1/3 \leq \mu_n^{\text{act}}(A_n) \leq 1/2$ for all sufficiently large n on the good event. A one-site update changes q_1 by at most

$$\Delta_n := \frac{2 \max_i |z_{1,n}(i)|}{\sqrt{n}} = O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n}\right)$$

by (6). Thus every one-step transition from A_n to A_n^c starts in

$$\{-\Delta_n \leq q_1 < 0\} \subseteq \{|q_1| \leq \varepsilon\}$$

for all sufficiently large n . The stationary flow therefore satisfies

$$Q(A_n, A_n^c) \leq e^{-c_1 n},$$

and the conductance of this set obeys

$$\Phi(A_n) := \frac{Q(A_n, A_n^c)}{\mu_n^{\text{act}}(A_n)} \leq 3e^{-c_1 n}.$$

For completeness, let $\pi = \mu_n^{\text{act}}$ and $\pi_{A_n} = \pi(\cdot \mid A_n)$. Reversibility gives

$$\|\pi_{A_n} P - \pi_{A_n}\|_{\text{TV}} = \Phi(A_n).$$

By contraction of total variation under a Markov kernel and a telescoping sum,

$$\|\pi_{A_n} P^t - \pi_{A_n}\|_{\text{TV}} \leq t\Phi(A_n).$$

Therefore

$$\|\pi_{A_n} P^t - \pi\|_{\text{TV}} \geq \pi_{A_n} P^t(A_n) - \pi(A_n) \geq \frac{1}{2} - t\Phi(A_n).$$

For $t \leq (4\Phi(A_n))^{-1}$, this distance is at least $1/4$. Since $\pi_{A_n} P^t$ is a convex combination of the laws $P^t(x, \cdot)$, $x \in A_n$, at least one point-mass initial state has distance at least $1/4$. This yields the stated exponential mixing-time lower bound, after decreasing c_1 if necessary. $\square$

8 Embedding by a finite-for-each- n disorder-dependent field

Let $S = S_n$ and $C = S^c$. Fix any $\sigma \in \{\pm 1\}^C$. Define the field on active coordinates by

$$h_i = -\beta \sum_{k \in C} W_{ik} \sigma_k, \quad i \in S, \quad (22)$$

and on pinned coordinates by

$$h_k = H_n \sigma_k, \quad k \in C, \quad H_n = n^2. \quad (23)$$

Thus the complement C is strongly pinned, while the field on S is chosen only to cancel the interaction generated by the pinned spins.

Lemma 8.1 (Pinning and exact active conditional law). *With probability tending to one,*

$$\mu_{W,h}(X_C \neq \sigma) \leq e^{-n^{3/2}}.$$

Conditional on $X_C = \sigma$, the law of X_S is exactly the zero-field active measure μ_n^{act} .

Proof. For every $k \in C$ and every configuration of the other spins, the signed local field satisfies

$$\sigma_k \left(h_k + \beta \sum_{j \neq k} W_{kj} X_j \right) \geq H_n - \beta \sum_{j \neq k} |W_{kj}|.$$

For one row of the GOE,

$$\mathbb{E} \sum_{j \neq k} |W_{kj}| = (n-1) \sqrt{\frac{2}{\pi n}} = O(\sqrt{n}).$$

Moreover, as a function of the underlying standard Gaussian row, the sum has Lipschitz constant at most one. Gaussian concentration followed by a union bound over k therefore gives

$$\max_k \sum_{j \neq k} |W_{kj}| = O_{\mathbb{P}}(\sqrt{n}).$$

On this event, if L_k denotes the preceding signed local field, then $L_k \geq n^2 - O(\sqrt{n})$ and the heat-bath conditional probability of the wrong sign satisfies

$$\mu_{W,h}(X_k \neq \sigma_k \mid X_{-k}) = \frac{1}{1 + e^{2L_k}} \leq e^{-2n^2 + O(\sqrt{n})}.$$

Taking expectation with respect to the Gibbs law and summing over $k \in C$ gives

$$\mu_{W,h}(X_C \neq \sigma) \leq ne^{-2n^2 + O(\sqrt{n})} \leq e^{-n^{3/2}}$$

for all sufficiently large n .

If $X_C = \sigma$, the cross interaction in the Hamiltonian is

$$\beta X_S^\top W_{S,C} \sigma,$$

whereas the active external-field contribution is

$$h_S^\top X_S = -\beta X_S^\top W_{S,C} \sigma.$$

They cancel exactly. All remaining terms involving only X_C are constant under the conditioning, and the X_S -dependent Hamiltonian is precisely

$$\frac{\beta}{2} X_S^\top W_{S,S} X_S.$$

Thus the conditional law of X_S is μ_n^{act} . □

Proof of Theorem 1.1. First work under the joint law of (W, q_0) . On an event of probability tending to one, Proposition 7.1 supplies an active bottleneck. Construct h by (22)–(23) and define

$$\tilde{A}_n = \{X_C = \sigma, q_1(X_S) < 0\}.$$

By Lemma 8.1 and the exact conditional active law,

$$\mu_{W,h}(\tilde{A}_n) = \mu_{W,h}(X_C = \sigma) \mu_n^{\text{act}}(A_n) = \frac{1}{2} + o(1).$$

Flow out of $\tilde{A}_n$ can occur in two ways. First, suppose an active coordinate is selected while $X_C = \sigma$. The conditional update kernel on S is exactly the active heat-bath kernel. Since the full chain chooses an active coordinate with probability $|S|/n$,

$$Q_{\text{full}}^{(S)}(\tilde{A}_n, \tilde{A}_n^c) = \mu_{W,h}(X_C = \sigma) \frac{|S|}{n} Q_{\text{act}}(A_n, A_n^c) \leq e^{-c_1 n}.$$

Second, if a pinned coordinate is selected, its probability of being resampled away from σ is uniformly at most $e^{-2n^2 + O(\sqrt{n})}$. Hence

$$Q_{\text{full}}^{(C)}(\tilde{A}_n, \tilde{A}_n^c) \leq e^{-n^{3/2}}.$$

Since $\mu_{W,h}(\tilde{A}_n)$ is bounded below by a positive constant,

$$\Phi(\tilde{A}_n) \leq e^{-cn}$$

for some deterministic $c > 0$. The same conductance-to-mixing argument used in Proposition 7.1 yields

$$t_{\text{mix}}(1/4; W, h) \geq e^{cn}$$

after adjusting c .

It remains to remove the auxiliary Haar vector q_0 . Let $\mathcal{G}_n(W, q_0)$ be the event that the construction succeeds, and define

$$g_n(W) = \mathbb{P}_{q_0}(\mathcal{G}_n(W, q_0) \mid W).$$

We have $0 \leq g_n \leq 1$ and

$$\mathbb{E}_W g_n(W) = \mathbb{P}_{W, q_0}(\mathcal{G}_n) \longrightarrow 1.$$

Therefore

$$\mathbb{P}_W(g_n(W) > 0) \geq \mathbb{E}_W g_n(W) \longrightarrow 1.$$

Whenever $g_n(W) > 0$, there exists at least one choice of q_0 , and hence at least one corresponding field $h \in \mathbb{R}^n$, for which the full chain has exponential mixing time. This proves the first assertion of the theorem without requiring a measurable choice $W \mapsto h(W)$. The polynomial consequence follows immediately because $e^{cn} > n^K$ for every fixed K and all sufficiently large n . $\square$

9 Interpretation for the TAP/non-backtracking program

The proof identifies the information missed by a source-only, pinning-uniform TAP argument. A finite spectral query extracts a path-like mode with eigenvalue close to 2. Thresholding its coordinates retains a disproportionately large amount of signal while reducing the Gaussian bulk dimension. The two relevant scalings are

$$\text{signal: } \ell(a)^2, \quad \text{annealed bulk cost: } p(a)^2.$$

For the explicit parameters above, the product-measure signal gain exceeds the entire annealed Gaussian free-energy contribution on the separating slab. Consequently, no valid refinement of a source-only local expansion can establish the false simultaneous-over-all-disorder-adaptive-fields statement under the present quantifiers.

The rigorous conclusion is therefore about quantifiers: a high-probability statement uniform over all fields chosen after observing the disorder cannot extend to every $\beta < 1$. A plausible remaining positive problem is the quenched statement for each field sequence fixed independently of the disorder.