hexagon:2610.00402v1

Lower and Upper Bounds for Glauber Dynamics from Typical States of the Low-Temperature Sherrington–Kirkpatrick Model

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

With the exception of this note, the entire document is machine-written. We have carefully supervised its preparation to ensure that the proofs are readable and that previous literature is appropriately cited. We are releasing this manuscript on Hexagon to disseminate the result as quickly as possible because it may be of interest to the mathematical community. We welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

P. L. was partially supported by NSF grant DMS-2450004.

ABSTRACT

We study the heat-bath (Glauber) dynamics of the Sherrington–Kirkpatrick model at zero external field and fixed inverse temperature $\beta > 1$. We sample the disorder J and then a configuration X_0 from the Gibbs measure, and run the continuous-time dynamics with rate one per site from X_0. Let $T^J(X_0)$ be the first time at which the law of the configuration, conditional on J and the realized X_0, is within total-variation distance $1/4$ of the Gibbs measure. We prove lower and upper bounds: for every fixed $\omega \in (0, 1/4)$, with probability tending to one,

$$A_\beta n^{7/30} < \log T^J(X_0) \le n^{1-\omega},$$

where $A_\beta > 0$ depends only on β. Hence, if $\log T^J(X_0) = n^{\zeta+o(1)}$, then $7/30 \le \zeta \le 3/4$; the physics literature suggests $\zeta = 1/3$. Both bounds also hold for the discrete-time chain.

These results improve on OpenAI's work in the same setting in two ways. First, OpenAI proved a lower bound with exponent $1/10000$ [77, Theorem 1.1] and an upper bound with exponent $1 - 1/40000000$ [85, Theorem 1]. We follow the architecture of both proofs. The lower bound adds a three-replica comparison whose pressure deficit is cubic in the width of the overlap window, together with the $n^{-7/10}$ upper bound of Aronow and Lopatto [6, Theorem 1.1] on the mean bias. The upper bound adds a pathwise cancellation between information and energy, concentration estimates on a growing mesh, and a choice of exponents that reaches every saving below $1/4$. Second, OpenAI formalized its lower bound in the Lean 4 proof assistant but not its upper bound; we have formalized both of our bounds in Lean.

We also formalize two known results that we present. The first is OpenAI's theorem that the variance of the logarithm of the partition function is $n^{1/3+o(1)}$. OpenAI's own formalization of this theorem assumes a bound on positive logarithmic moments that is proved in their paper but not formalized; our formal proof establishes this bound. The second is the consequence, recorded by Aronow and Lopatto [6, Section 1.6], that the mean bias of the pressure is at least of order $n^{-5/6-\varepsilon}$ for every fixed $\varepsilon > 0$.

Contents

1 Introduction 1
- 1.1 Main results . . . 2
- 1.2 Two known results: fluctuations and the mean bias . . . 5
- 1.3 Related results on the dynamics. . . . 6
- 1.4 The fluctuation scale and the mean bias . . . 8
- 1.5 Formal verification . . . 11
- 1.6 Methods and their sources . . . 12
- 1.7 Organization of the paper. . . . 20

2 Overview of the proofs 22
- 2.1 The sign test . . . 22
- 2.2 Overlap constraints and reference configurations . . . 24
- 2.3 The exponent . . . 26
- 2.4 The cubic deficit . . . 28
- 2.5 From pressures to Gibbs probabilities . . . 31
- 2.6 The dynamical argument . . . 33
- 2.7 The upper bound . . . 35
- 2.8 Fluctuations and the mean bias . . . 45

PART I Preliminaries 47

3 The model and the dynamics 48
- 3.1 The Sherrington–Kirkpatrick model . . . 48
- 3.2 The completed Hamiltonian . . . 50
- 3.3 Heat-bath dynamics . . . 52
- 3.4 The realized-start experiment . . . 55
- 3.5 Asymptotic conventions . . . 57
- 3.6 Normalizations in the cited works . . . 58

4 Gaussian hierarchies and imported results 61
- 4.1 Gaussian integration steps . . . 62
- 4.2 The Parisi functional and the scalar diffusion . . . 64
- 4.3 Finite-size estimates . . . 68
- 4.4 Gaussian hierarchies and constrained pressures . . . 70
- 4.5 Two- and three-replica pressure bounds . . . 77

4.6 Energy, disorder rotation, and overlap-neighborhood masses 80

PART II Static estimates for the lower bound 82

5 The scalar Parisi problem on the support 83
5.1 Bounds on spatial derivatives 84
5.2 Stochastic identities along the scalar diffusion 88
5.3 The Parisi distribution function on $(0, q_\beta)$ 91

6 Short-time estimates for Gaussian hierarchies 95
6.1 Flows and the main estimates 96
6.2 Three tools 101
6.3 Spatial derivatives and the step limit 105
6.4 Parameter derivatives 110
6.5 Response to a change of the terminal function 113

7 A three-replica covariance perturbation 117
7.1 The perturbed covariance path 119
7.2 Derivatives of the local block 126
7.3 Taylor coefficients at constant mass 128
7.4 The varying profile 138
7.5 The expansion of the trial value 142
7.6 A uniform cubic deficit on a band 143

8 Fractional moments of constrained partition functions 147
8.1 Hierarchies with a positive outer mass 148
8.2 The fractional matrix comparison 150
8.3 The profile $\max\{\alpha, 3s\}$ 153
8.4 A change of profile before the local window 153
8.5 The local fractional bound 157

9 A simultaneous estimate for overlap constraints 161
9.1 Orientations, pairs and the fixed scale 162
9.2 Separated overlaps 163
9.3 Families of constraints 170
9.4 The simultaneous event 172

PART III Dynamics: proof of the lower bound 177

10 Overlap neighborhoods and reference configurations 178
10.1 Overlap neighborhoods at sublinear scales 179
10.2 Parameters and exceptional triples 183
10.3 Fixed families of reference configurations 189

11 Stable overlap constraints along a trajectory 194
11.1 Standing assumptions and the threshold geometry . . . 195
11.2 The selection and crossing algorithm . . . 198
11.3 One level of the algorithm . . . 200
11.4 The induction . . . 205

12 Proof of the barrier theorem 207
12.1 Predictable thresholds . . . 208
12.2 A deterministic sign obstruction . . . 209
12.3 The choice of the coefficient . . . 210
12.4 Stationarity, the clocks, and total variation . . . 210
12.5 The parameters and the exponents . . . 213
12.6 Proof of Theorem 1.1 . . . 213
12.7 Quantitative and almost-sure forms . . . 215

PART IV The upper bound 224

13 Primitive disorder, the scalar process beyond the support, and the parameters 225
13.1 Primitive disorder and the completed Hamiltonian . . . 225
13.2 Gaussian orthogonal ensemble facts . . . 227
13.3 The exponent parameters . . . 229
13.4 The scalar Parisi process and its continuation . . . 229

14 The Gaussian-query proposal on a growing mesh 235
14.1 Growing schedule and source law . . . 235
14.2 Spin laws, histories, and filtrations . . . 238
14.3 Population entries . . . 241

15 Forward entropy at the fine horizon 244
15.1 The reference measure . . . 244
15.2 Fine energy . . . 247

16 Concentration on the growing mesh 252
16.1 Errors and dyadic depth . . . 252
16.2 Weighted dyadic concentration . . . 254

17 The joint cancellation and coarse domination 260
17.1 Dyadic algebra and the row telescope . . . 260
17.2 The joint information–energy cancellation . . . 264
17.3 Coarse domination . . . 271

18 Smoothing, coverage, and the connecting sequence 275
18.1 Smoothing and coverage . . . 275
18.2 The witness event and the enlarged boundary . . . 278

18.3 The connecting sequence . . . 280
18.4 Exclusion of low-flow witnesses. . . . 285

19 Proof of the upper bound and the parameter system 287
19.1 From balanced-cut flow to mixing. . . . 287
19.2 Proof of the main theorem . . . 290
19.3 Proof of the discrete-time corollary . . . 291
19.4 Proof of the two-sided corollary. . . . 292
19.5 The parameter system . . . 293

PART V Fluctuations and mean bias 297

20 The fluctuation scale 298
20.1 The two-sided bound on positive logarithmic moments . . . 298
20.2 OpenAI's proof of the two-sided bound . . . 299
20.3 Log-concave augmentation . . . 305
20.4 Proof of the fluctuation theorem . . . 308

21 The finite-size mean bias 311
21.1 The variance–bias inequality . . . 311
21.2 Proof of the mean-bias bound. . . . 312
21.3 What the bound does not determine . . . 313

PART VI Formal verification and open problems 314

22 Formal verification 315
22.1 OpenAI's formalization. . . . 315
22.2 Statements of the formal results. . . . 316
22.3 The formal and the written barrier proofs . . . 318
22.4 The formal proof of the two-sided bound . . . 319
22.5 The upper bound . . . 321
22.6 Scope and availability. . . . 322

23 Open problems 324

References 327

SECTION 1

Introduction

The Sherrington–Kirkpatrick (SK) model is the mean-field Ising spin glass, in which every pair of spins interacts through an independent Gaussian coupling [100]. Its equilibrium theory is well developed. Parisi's hierarchical ansatz [91] predicted the limiting free energy, and this prediction was proved through Guerra's interpolation bound [59] and Talagrand's matching lower bound [103]. At every fixed inverse temperature $\beta > 1$ and zero external field, the minimizing Parisi measure is supported on an interval $[0, q_\beta]$; this was proved by Zhou for β close to one [116, Theorem 1] and in [69, Theorem 1.1] for every $\beta > 1$. Much less is known about the dynamics of the model at low temperature, and about the finite-size corrections to its equilibrium quantities. This paper addresses both.

Our main results concern the heat-bath dynamics, also called Glauber dynamics [58], in which one spin at a time is resampled from its conditional law. For fixed $\beta > 1$, we sample the couplings and then one configuration from the Gibbs measure, and run the dynamics from this configuration. We prove that, with probability tending to one, the law of the configuration at time $\exp(A_\beta n^{7/30})$ is still at total-variation distance greater than τ from the Gibbs measure, for every fixed $\tau < 1/2$ and a coefficient $A_\beta > 0$ that does not depend on τ (Theorem 1.1). In the other direction, for every fixed $\omega \in (0, 1/4)$, the distance is at most $1/4$ at time $\exp(n^{1-\omega})$, again with probability tending to one (Theorem 1.3). OpenAI proved bounds of both kinds in the same setting, with the same joint law of the couplings and the initial configuration, the same clocks of rate one per site, and the threshold $1/4$: a lower bound at time $\exp(n^{1/10000})$ [77, Theorem 1.1] and an upper bound at time $\exp(n^{1-1/40000000})$ [85, Theorem 1]. We follow the overall strategy of both proofs. Let $T^J(X_0)$ be the first time at which the distance from the Gibbs measure is at most $1/4$. Combining our two bounds, we find that if $\log T^J(X_0) = n^{\zeta+o(1)}$, then $7/30 \le \zeta \le 3/4$ (Corollary 1.5). The proof of the lower bound uses the upper bound $n^{-7/10}$ of Aronow and Lopatto on the mean bias of the pressure [6, Theorem 1.1].

Our second contribution is a complete formal proof of Theorem 1.1 in the Lean 4 proof assistant [48, 108], together with formal proofs of two results that are not new. The first is the fluctuation theorem $\mathrm{Var}(F_n) = n^{1/3+o(1)}$ for the logarithm F_n of the partition function (Theorem 1.6), which was stated and proved by OpenAI in [81, Theorem 1.1]. OpenAI's public formalization of this theorem assumes a two-sided bound on positive logarithmic moments of the centered completed log partition function, which is proved

on paper in [81, Proposition 7.5]; our formal proof establishes this bound. We restate the theorem with attribution, describe the structure of its proof, and reproduce the short final step of that proof; we do not claim a new mathematical proof of Theorem 1.6. We have formalized Theorem 1.3; see Section 22.5.

The second known result is a lower bound with exponent $5/6$ on the mean bias $P_\beta - n^{-1}\mathbb{E}F_n$, where $P_\beta = \lim_{n\to\infty} n^{-1}\mathbb{E}F_n$ is the limiting pressure. It follows from Theorem 1.6 and a variance–bias inequality of Aronow and Lopatto [8, Lemma 7.4 and Remark 7.14] that the mean bias is at least of order $n^{-5/6-\varepsilon}$ for every fixed $\varepsilon > 0$ (Corollary 1.7). This deduction was already recorded in [6, Section 1.6]; what is new here is its formal proof. Together with the upper bound of [6], it shows that if the mean bias is $n^{-a+o(1)}$, then $7/10 \le a \le 5/6$.

1.1 Main results

For an integer $n \ge 2$, let $\Sigma_n = \{-1,1\}^n$, and let $J = (J_{ij})_{1\le i<j\le n}$ be independent standard Gaussian variables. We set $J_{ji} = J_{ij}$ and $J_{ii} = 0$. The Hamiltonian, the partition function, and the Gibbs measure at inverse temperature β are

$$H_n(x) = \frac{1}{\sqrt{n}} \sum_{1\le i<j\le n} J_{ij}x_ix_j, \qquad Z_n = \sum_{x\in\Sigma_n} e^{\beta H_n(x)}, \qquad \pi^J(x) = \frac{e^{\beta H_n(x)}}{Z_n}. \tag{1.1}$$

We write $R(x,y) = n^{-1}\sum_{i=1}^n x_iy_i$ for the overlap of $x, y \in \Sigma_n$, and $F_n = \log Z_n$.

In the heat-bath dynamics, each site carries an independent exponential clock of rate one. When the clock of site i rings at a configuration x, the spin x_i is replaced by $\sigma \in \{-1,1\}$ with probability

$$p_i^J(\sigma \mid x) = \frac{e^{\beta\sigma h_i(x)}}{2\cosh(\beta h_i(x))}, \qquad h_i(x) = \frac{1}{\sqrt{n}} \sum_{j\ne i} J_{ij}x_j,$$

which is the conditional law of the ith spin under π^J given the other spins. An update may leave the configuration unchanged. We write $P_t^J(x,\cdot)$ for the law at time t of this process started at x, and K^J for the transition matrix of the discrete chain that chooses a site uniformly at random and performs one such update. One unit of continuous time contains n update attempts on average. Both chains are reversible with respect to π^J. For $x \in \Sigma_n$, $t \ge 0$, and integers $k \ge 0$, we set

$$d^J(x,t) = \|P_t^J(x,\cdot) - \pi^J\|_{\mathrm{TV}}, \qquad d^J_{\mathrm{disc}}(x,k) = \|(K^J)^k(x,\cdot) - \pi^J\|_{\mathrm{TV}},$$

where $\|\nu - \zeta\|_{\mathrm{TV}} = \frac{1}{2}\sum_x |\nu(x) - \zeta(x)|$. We also write

$$T^J(x) = \inf\{t \ge 0 : d^J(x,t) \le 1/4\}, \qquad T^J_{\mathrm{disc}}(x) = \min\{k \ge 0 : d^J_{\mathrm{disc}}(x,k) \le 1/4\}$$

for the mixing times of the two chains started at x. Since $t \mapsto d^J(x,t)$ is continuous and nonincreasing (Lemma 3.2), we have $T^J(x) \le t$ if and only if $d^J(x,t) \le 1/4$, and similarly in discrete time. Finally, $\mathbb{P}_{J,\,X_0\sim\pi^J}$ denotes the joint law of the couplings J and of a configuration X_0 whose conditional law given J is π^J; see (3.16). These definitions are given in detail in Section 3.

Theorem 1.1 (Barrier from typical equilibrium states)**.** *For every fixed $\beta > 1$ there exists $A_\beta > 0$ such that, for every fixed $\tau \in (0, 1/2)$,*

$$\mathbb{P}_{J,\,X_0\sim\pi^J}\{d^J(X_0, e^{A_\beta n^{7/30}}) > \tau\} \longrightarrow 1, \tag{1.2}$$

$$\mathbb{P}_{J,\,X_0\sim\pi^J}\{d^J_{\mathrm{disc}}(X_0, \lfloor e^{A_\beta n^{7/30}} \rfloor) > \tau\} \longrightarrow 1. \tag{1.3}$$

The configuration X_0 is sampled once and then held fixed, so that $d^J(X_0,t)$ measures the distance from equilibrium of the chain started at this particular configuration. If the initial configuration were instead averaged over π^J before the distance is taken, the law at time t would be π^J by stationarity, and the distance would vanish. Since the Gibbs mass of the set of initial configurations in the event of (1.2) takes values in $[0,1]$ and has expectation equal to the displayed probability, (1.2) is equivalent to the convergence of this Gibbs mass to one in probability over the disorder; this observation is made after [77, Theorem 1.1] and is recalled in Section 3. The coefficient A_β depends only on β: the same coefficient serves for every fixed threshold $\tau < 1/2$ and for both clocks, while the rate of convergence in (1.2) and (1.3) may depend on τ. The restriction $\tau < 1/2$ comes from spin-flip symmetry: for every configuration v, the half-space $\{y : R(v,y) \le 0\}$ has Gibbs mass at least $1/2$, and the proof shows that, for a suitable configuration v with $R(v, X_0)$ bounded away from zero, the chain started at X_0 is unlikely to enter this half-space before the stated time.

The exponent $7/30$ is one third of the exponent $7/10$ in the bound $0 \le P_\beta - n^{-1}\mathbb{E}F_n \le C_b n^{-7/10}$ on the mean bias proved in [6, Theorem 1.1], where we recall that P_β is the limiting pressure. The proof needs overlap constraints on a scale ρ whose pressure deficit, of order ρ^3 per spin, dominates the mean bias, and our counting argument allows a time t with $\log t$ of order $1/\rho$. With the bound $n^{-7/10}$, we can take $\rho \asymp n^{-7/30}$. Section 1.6 explains this balance in more detail. A union bound over polynomially many constraints in the argument requires $n\rho^4$ to be large compared with $\log n$, so the present argument gives no exponent of $1/4$ or more, whatever the upper bound on the mean bias. Theorem 1.1 is proved in Section 12.

Since total variation contracts under a Markov kernel that preserves π^J, the distances $d^J(x,t)$ and $d^J_{\mathrm{disc}}(x,k)$ are nonincreasing in t and k. For every fixed $A > 0$ and $\kappa < 7/30$, we have $n^\kappa < An^{7/30}$ for all sufficiently large n. Theorem 1.1 implies the following statement with coefficient one, which is also proved in Section 12.

Corollary 1.2 (Unit coefficient below the exponent $7/30$)**.** *For every fixed $\beta > 1$, every fixed $\kappa \in (0, 7/30)$, and every fixed $\tau \in (0, 1/2)$,*

$$\mathbb{P}_{J,\, X_0 \sim \pi^J}\{d^J(X_0, e^{n^\kappa}) > \tau\} \longrightarrow 1, \qquad \mathbb{P}_{J,\, X_0 \sim \pi^J}\{d^J_{\mathrm{disc}}(X_0, \lfloor e^{n^\kappa} \rfloor) > \tau\} \longrightarrow 1.$$

The corollary does not assert coefficient one at the exponent $7/30$ itself.

The proof gives more than the convergence of the probabilities in Theorem 1.1; we record the stronger forms in Theorem 12.6. There exist $A > 0$ and, for each n, a disorder event of probability at least $1 - Ce^{-cn^{1/15}}$ on which the following holds for all sufficiently large n. The set of initial configurations x with $d^J(x, t) \geq 1/2 - Ce^{-cn^{7/30}}$ for every $t \leq e^{An^{7/30}}$, and with the analogous bound for the discrete chain, has Gibbs mass at least $1 - Ce^{-cn^{7/30}}$. The same holds for the Metropolis dynamics and, more generally, for every single-site dynamics with uniformly bounded rates that leaves π^J stationary. For every reversible such dynamics, the relaxation time is at least $e^{An^{7/30}}/(Cn^{7/30})$. Finally, the Gibbs mass of the slow initial configurations tends to one almost surely along the full sequence of system sizes, however the disorders of different sizes are coupled. The main change in the proof is a lower bound on F_n from the small-deviation inequality of Paouris and Valettas for convex functions of Gaussian vectors [90, Theorem 2.1 and Remark 2.4(3)], in place of Chebyshev's inequality. We have formalized Theorem 1.1 and Corollary 1.2 in Lean, but not Theorem 12.6; see Section 1.5.

Theorem 1.3 (Mixing from typical equilibrium states)**.** *For every fixed $\beta > 1$ and every fixed $\omega \in (0, 1/4)$,*

$$\mathbb{P}_{J,\, X_0 \sim \pi^J}\{d^J(X_0, e^{n^{1-\omega}}) \leq 1/4\} \longrightarrow 1; \quad \textit{equivalently} \quad \mathbb{P}_{J,\, X_0 \sim \pi^J}\{T^J(X_0) \leq e^{n^{1-\omega}}\} \longrightarrow 1.$$

As in Theorem 1.1, the configuration X_0 is sampled once and then held fixed when the distance $d^J(X_0, t)$ is formed, and the clocks have rate one per site, as in Section 3.3. By the argument of Lemma 3.4, applied to the complementary events $\{d^J(X_0, t) > 1/4\}$, Theorem 1.3 is equivalent to the convergence to zero, in probability over the disorder, of the Gibbs mass of the initial configurations x with $d^J(x, e^{n^{1-\omega}}) > 1/4$. The theorem asserts nothing at $\omega = 1/4$. It improves the exponent $1 - 1/40000000$ of [85, Theorem 1], and it is proved in Part IV. We have formalized Theorem 1.3; see Section 22.5.

For the discrete chain, we prove the same bound after $\lfloor e^{n^{1-\omega}} \rfloor$ update attempts. Since one unit of continuous time contains n update attempts on average, this number of attempts corresponds to the time $e^{n^{1-\omega}}/n$. Together with Theorem 1.1, we obtain two-sided bounds in both time conventions.

Corollary 1.4 (Discrete time)**.** *For every fixed $\beta > 1$ and $\omega \in (0, 1/4)$,*

$$\mathbb{P}_{J, X_0 \sim \pi^J}\{d^J_{\mathrm{disc}}(X_0, \lfloor e^{n^{1-\omega}} \rfloor) \leq 1/4\} \to 1.$$

Corollary 1.5 (Two-sided typical-start bounds)**.** *For every fixed $\beta > 1$ and $\omega \in (0, 1/4)$, with A_β from Theorem 1.1,*

$$\mathbb{P}_{J, X_0 \sim \pi^J}\{e^{A_\beta n^{7/30}} < T^J(X_0) \le e^{n^{1-\omega}}\} \to 1,$$
$$\mathbb{P}_{J, X_0 \sim \pi^J}\{\lfloor e^{A_\beta n^{7/30}} \rfloor < T^J_{\mathrm{disc}}(X_0) \le \lfloor e^{n^{1-\omega}} \rfloor\} \to 1.$$

In the notation of Section 3.5, Corollary 1.5 implies that if $\log T^J(X_0) = n^{\zeta + o_{\mathbb{P}}(1)}$ under $\mathbb{P}_{J, X_0 \sim \pi^J}$ for some ζ, then $7/30 \le \zeta \le 3/4$, and the same holds for $T^J_{\mathrm{disc}}(X_0)$. The value $\zeta = 1/3$ suggested by the physics literature (Section 1.3) lies in this interval. Corollaries 1.4 and 1.5 are proved in Section 19.

Both arguments stop at the exponent $1/4$. The proof of Theorem 1.1 needs $n\rho^4/\log n \to \infty$ for the overlap scale ρ, and the proof of Theorem 1.3 needs the chain of inequalities $\omega < \mathsf{q} < \mathsf{s_b} < \mathsf{s_{seq}} < \mathsf{e} < \mathsf{u_E} < \mathsf{k_f}/2 \le 1/4$ of Section 19.5. We do not know whether this coincidence has a common explanation.

1.2 Two known results: fluctuations and the mean bias

The other two results that we present concern the fluctuations of F_n and the mean bias. The first was proved by OpenAI. The paper [81] uses the uniform probability measure on Σ_n instead of the counting measure in (1.1). This changes F_n by the constant $n \log 2$ and leaves the following statement unchanged (see Section 3.6).

Theorem 1.6 ([81, Theorem 1.1])**.** *For every fixed $\beta > 1$ and every $\eta > 0$, there exists $n_0(\beta, \eta)$ such that, for all $n \ge n_0$,*

$$n^{1/6-\eta} \le \sqrt{\mathrm{Var}(F_n)} \le n^{1/6+\eta}, \tag{1.4}$$
$$n^{1/3-\eta} \le \mathrm{Var}(F_n) \le n^{1/3+\eta}. \tag{1.5}$$

Moreover, for every $\eta > 0$,

$$\mathbb{P}\{n^{1/6-\eta} \le |F_n - \mathbb{E}F_n| \le n^{1/6+\eta}\} \longrightarrow 1. \tag{1.6}$$

The proof of Theorem 1.6 in [81] reduces it to a two-sided bound on positive logarithmic moments of the centered completed log partition function, which we state as Proposition 20.1 and discuss in Section 1.4. Our formal proof of Theorem 1.6 includes a formal proof of this bound (Section 1.5).

Let $p_n(\beta) = n^{-1}\mathbb{E}F_n$, so that the mean bias is $\Delta_{n,\beta} = P_\beta - p_n(\beta)$. By Guerra's bound [59, Theorem 3] and the Parisi formula [103, Theorem 1.1], we have $\Delta_{n,\beta} \ge 0$ for every n. A lower bound on the variance implies a lower bound on the mean bias. Aronow and Lopatto proved that for every fixed $\beta > 1$ there exists $c_\beta > 0$ such that

$$\Delta_{n,\beta} \ge c_\beta \left(\frac{\mathrm{Var}(F_n) + \beta^2/2}{n}\right)^{5/4} \tag{1.7}$$

for every $n \geq 2$ [8, Lemma 7.4 and Remark 7.14, equation (7.68)]; we state it as Proposition 21.1. The term $\beta^2/2$ is the variance of the Gaussian variable that completes the covariance of βH_n (Section 3.2). Inserting (1.5) into (1.7), we obtain the following corollary, which is already recorded in [6, Section 1.6, p. 11].

Corollary 1.7 (Mean bias)**.** *For every fixed $\beta > 1$ and every $\varepsilon > 0$, there exist $c_{\beta,\varepsilon} > 0$ and $n_0(\beta, \varepsilon)$ such that*

$$\Delta_{n,\beta} \geq c_{\beta,\varepsilon} n^{-5/6-\varepsilon} \qquad (n \geq n_0(\beta, \varepsilon)). \tag{1.8}$$

Equivalently, $nP_\beta - \mathbb{E}F_n \geq c_{\beta,\varepsilon} n^{1/6-\varepsilon}$ for $n \geq n_0(\beta, \varepsilon)$.

Together with the upper bound of [6], Corollary 1.7 implies that if $\Delta_{n,\beta} = n^{-a+o(1)}$ for some a, then $7/10 \leq a \leq 5/6$. Section 21 proves Corollary 1.7, which does not use Theorem 1.1, and discusses what it leaves open. If $\liminf_{n\to\infty} n^{-1/3} \operatorname{Var}(F_n) > 0$, the same inequality implies that $\Delta_{n,\beta} \geq c_\beta n^{-5/6}$ for all sufficiently large n, as observed in [6, Section 1.6, equation (1.4)]; this hypothesis follows from the convergence of $n^{-1/3} \operatorname{Var}(F_n)$ to a positive constant, which is stated in [82, Theorem 1.2]. Our formal proof of Corollary 1.7 combines the formal proof of Theorem 1.6 with a formal proof of (1.7).

1.3 Related results on the dynamics

Physics. The slow relaxation of finite SK systems has been studied in physics since the early 1980s. Sompolinsky and Zippelius developed the dynamical mean-field theory of the spin-glass phase [101, 102], in which the relaxation times of a finite system diverge with its size n. In simulations of the Ising SK model, Mackenzie and Young reported two families of time scales τ_n. For observables invariant under global spin reversal, the behavior of $\log \tau_n$ was consistent with $n^{1/4}$; for the reversal itself, the disorder average of $\log \tau_n$ was consistent with $n^{1/2}$ [71, 72]. Later work revised this picture. Vertechi and Virasoro computed energy barriers between metastable states numerically and found growth like n^α with $\alpha = 0.34 \pm 0.08$ [114]. Colborne's Monte Carlo study gave the exponent 0.25 ± 0.05 for the barrier that governs the ergodic time [40]. Kinzelbach and Horner predicted free-energy barriers of order $n^{1/3}$ from a resummed perturbative analysis of the dynamics of finite systems [65]. The simulations of Billoire and Marinari [19], of Bittner and Janke [20], and of Billoire [18] are consistent with $\log \tau_n \propto n^{1/3}$, and so is the computation of spectral gaps of small systems by Monthus and Garel [76]. Aspelmeier and Moore studied barriers between solutions of the Thouless–Anderson–Palmer equations [31, 109]. They found that barriers of order $n^{1/3}$ are one of several possibilities below a threshold for the free-energy density [9].

These works concern different quantities: energy barriers, free-energy barriers along an overlap coordinate, autocorrelation times, and spectral gaps. They use binary as well as Gaussian couplings, Metropolis as well as heat-bath updates, and moderate system

sizes. They suggest that the logarithm of the relaxation time from a typical configuration grows like $n^{1/3}$, but they do not determine a mixing exponent. A different question concerns times that stay bounded as n grows. When $n \to \infty$ is taken first, the dynamics from a random configuration ages and does not reach equilibrium [46]; we refer to [26, Section 3.7] for a discussion of the two orders of limits.

High and critical temperature. Rigorous results on the dynamics depend strongly on the temperature. At high temperature, functional inequalities imply fast mixing from every initial configuration. Bauerschmidt and Bodineau proved a logarithmic Sobolev inequality, uniform in n, for a Dirichlet form of gradient type under a condition on the spectral width of the interaction matrix, which applies to SK when $\beta < 1/4$ [15, Theorem 1 and Corollary 2]. Eldan, Koehler, and Zeitouni proved a Poincaré inequality for the heat-bath Dirichlet form itself under a spectral condition of the same kind, and deduced a polynomial bound on the mixing time from every initial configuration [54, Theorems 1 and 11]. Anari, Jain, Koehler, Pham, and Vuong obtained mixing in $O(n \log n)$ updates in the same range through entropic independence [3, Theorem 12]. Anari, Koehler, and Vuong extended this range to a threshold of approximately 0.295 using localization schemes [4, Theorem 1 and Section 1.3]. Wang extended mixing in $O(n \log n)$ updates to every $\beta < 1/2$ [115, Theorem 1.1], and Boban, Li, and Oveis Gharan proved polynomial mixing for $\beta < 1/2 + \varepsilon_0$, for an absolute constant $\varepsilon_0 > 0$ [21, Theorem 1.5 and Corollary 1.6]. With external fields sufficiently large at every site and their average bounded by a fixed constant H, Paliy and Sohn proved a modified logarithmic Sobolev inequality and mixing in $O_{\beta,H}(n \log n)$ update attempts at every fixed β, with probability tending to one [87, Corollary 1.5]. For every fixed $\beta < 1$, OpenAI proved that the spectral gap of the continuous-time chain is bounded below by a positive constant with probability tending to one [84, Theorem 1.1], and that the chain started from the worst initial configuration exhibits cutoff at time $(\log n)/(2\lambda(\beta))$, for a deterministic $\lambda(\beta) > 0$ [80, Theorem 1.1]. At the critical temperature $\beta = 1$, OpenAI proved that, for every deterministic t_n with $t_n = o(n^{2/3})$, the Gibbs mass of initial configurations x with $d^J(x, t_n) > 1/4$ tends to one in probability [79, Theorem 1.1], and that the mixing time is $n^{2/3+o(1)}$ [78, Theorem 1.1].

Low temperature. At low temperature, Ben Arous and Jagannath showed that a free-energy barrier in the overlap of two replicas forces an exponentially small spectral gap in mean-field spin glasses [16, Theorem 1.3]. Their sufficient conditions for such a barrier include the vanishing of the second derivative at zero of the covariance function [16, Theorem 1.9], so they do not apply to the quadratic covariance of SK. For the pure spherical p-spin models with $p \geq 3$, Gheissari and Jagannath proved that the spectral gap of the Langevin dynamics is exponentially small at sufficiently low temperature [57, Theorem 1]. For SK, Sellke proved that there exist absolute constants $\beta_*, c > 0$ such that,

for every $\beta > \beta_*$, the mixing time from the worst initial configuration is at least $e^{c\beta n}$ with probability tending to one [99, Theorem 1.1]. The slow initial configurations in his proof are configurations at which almost every spin is aligned with its local field by a uniform margin. He notes after his theorem that the chain may mix much faster from a random initial configuration. El Alaoui, Montanari, and Sellke proved that for $\beta > 1$ no family of sampling algorithms that is stable under small perturbations of the disorder approximates the Gibbs measure in normalized Wasserstein distance [53, Theorem 2.6]. Neither result determines the Gibbs mass of the initial configurations from which the heat-bath chain mixes slowly.

OpenAI proved a lower bound and an upper bound for initial configurations sampled from the Gibbs measure that hold throughout $\beta > 1$. By [77, Theorem 1.1], the conclusions of Theorem 1.1 hold with the time $\exp(n^{1/10000})$ and the threshold $\tau = 1/4$. In a companion preprint, OpenAI proved that $\mathbb{P}_{J,\, X_0 \sim \pi^J}\{T^J(X_0) \le \exp(n^{1-c_\beta})\} \to 1$ for some $c_\beta \in (0,1)$, and the proof permits $c_\beta = 1/40000000$ [85, Theorem 1]. Combining the two bounds, and using that $t \mapsto d^J(x,t)$ is continuous and nonincreasing, they obtained a two-sided bound on $T^J(X_0)$ with these exponents [85, Corollary 2]. The upper bound of [85] is stated only in continuous time.

Theorem 1.1 improves the exponent $1/10000$ of [77] to $7/30$, and it holds for every fixed threshold below $1/2$ with a coefficient that does not depend on the threshold. Theorem 1.3 improves the exponent $1 - 1/40000000$ of [85] to $1 - \omega$ for every fixed $\omega < 1/4$, and Corollary 1.4 gives the same bound for the discrete chain. For $\beta > \beta_*$, we find from Theorem 1.3 that the mixing time from a typical initial configuration is much smaller than the lower bound $e^{c\beta n}$ of [99] on the mixing time from the worst initial configuration. Our results do not determine the exponent of the typical mixing time, which by Corollary 1.5 lies in $[7/30, 3/4]$ if it exists. They are compatible with the value $1/3$ suggested by the physics literature, but they do not decide whether this value is correct. The value $1/3$ also appears as the exponent of $\mathrm{Var}(F_n)$ in Theorem 1.6; we know of no argument relating the two exponents. The coefficient A_β depends on β, and our proof does not control it as $\beta \downarrow 1$ or $\beta \to \infty$. Similarly, the rate of convergence in Theorem 1.3 depends on β and ω, and we assert no uniformity in either parameter.

1.4 The fluctuation scale and the mean bias

For a survey of free-energy fluctuations in the SK and spherical SK models, including their multi-species variants and the effects of an external field, see Collins-Woodfin and Le [41].

Physics. The prediction of the fluctuation scale of F_n has a separate history in physics. In a truncated replica theory near the critical temperature, Kondor found that the replica-normalized free energy depends on a small positive replica number s first at order s^5 [66].

Crisanti, Paladin, Sommers, and Vulpiani interpreted the replica number as a tilt of the disorder distribution, and predicted that the free energy per spin fluctuates on the scale $n^{-5/6}$ in the low-temperature phase at zero field [45]. At fixed β, this is the scale $n^{1/6}$ for F_n. Parisi and Rizzo extended the fifth-order calculation throughout the low-temperature phase and computed the resulting large deviations of the free energy [92, 93]. Passing from these large deviations to the width of typical fluctuations requires matching the tail of the distribution to its center. This is an additional assumption, which they discuss in [93, Section VI]; the limits $n \to \infty$ and $s \downarrow 0$ cannot simply be interchanged. Aspelmeier, Billoire, Marinari, and Moore argued that a system of n spins resolves about $n^{1/6}$ levels of replica symmetry breaking. From this they derived both a variance of order $n^{1/3}$ for the total free energy and a correction of order $n^{-2/3}$ to its mean per spin [11, Section 2]. Aspelmeier derived exact identities relating the variance of the free energy to overlaps between systems with correlated disorder [10]. At zero temperature, numerical studies of the width of the distribution of the ground-state energy have produced exponents between $3/4$ and $5/6$ for the standard deviation per spin, and the fits are sensitive to corrections to scaling [22, 23, 27, 86].

Rigorous results. Rigorous results on the fluctuations of F_n depend on the temperature and the field. For $0 < \beta < 1$ at zero field, F_n has Gaussian fluctuations of order one, as shown by Aizenman, Lebowitz, and Ruelle [1]. Comets and Neveu studied these fluctuations by stochastic calculus [42]. At $\beta = 1$ the variance grows logarithmically. Chen and Lam proved an upper bound of order $(\log n)^2$ [38, Theorem 1.1], Schertzer proved the upper bound $(\log n)/4 + C$ and a lower bound of order $\log\log\log n$ [98, Theorems 1 and 2], Tierz bounded the variance for every n in terms of the spin-glass susceptibility [110, Theorem 1.2], and Du and Huang proved that $\operatorname{Var}(F_n) = (\log n)/6 + O(1)$ and that the fluctuations have a Gaussian limit [52, Theorem 1.3]. Dey and Kang obtained the variance asymptotics and a Gaussian limit along sequences $\beta_n \uparrow 1$ for which $n^{1/3}(1-\beta_n^2)$ converges to a limit in $(0, \infty]$ [49, Theorem 1.1]. With a fixed nonzero external field, Chen, Dey, and Panchenko proved that the variance is of order n and that a central limit theorem holds [39, Theorems 1 and 2]. Dey and Wu identify a crossover at field strength $n^{-1/4}$ in the high-temperature regime [50, Theorem 1.3]. The spin space also matters: for the spherical SK model at low temperature, Baik and Lee proved that the free energy per spin has Tracy–Widom fluctuations of order $n^{-2/3}$ [14, Theorems 2.10 and 2.11], so that the total free energy fluctuates on the scale $n^{1/3}$ rather than $n^{1/6}$.

For fixed $\beta > 1$ at zero field, the Gaussian Poincaré inequality implies that $\operatorname{Var}(F_n) \le C_\beta n$. Chatterjee improved this to $O_\beta(n/\log n)$ and related this superconcentration to disorder chaos [33, Theorem 1.5]. He also proved that the fluctuations of F_n are at least of constant order at every fixed $\beta > 0$ and zero field, in the sense of nonconcentration on sufficiently short intervals [34, Theorem 2.4]. Aronow and Lopatto proved $c_\beta n^{4/15} \le \operatorname{Var}(F_n) \le C_v n^{7/15}$ for all sufficiently large n [8, Theorem 1.4], using

quantitative comparisons between positive fractional moments of Z_n and constrained Parisi functionals [8, Theorem 1.1]. The sharp exponent was then obtained by OpenAI [81, Theorem 1.1], as stated in Theorem 1.6.

Theorem 1.6 is a statement about exponents. It identifies neither a limiting law nor a positive limit of $n^{-1/3}\operatorname{Var}(F_n)$. Both are stated in the companion preprint [82, Theorems 1.1 and 1.2]: there, $n^{-1/3}\operatorname{Var}(F_n)$ converges to a constant $a_2(\beta) \in (0,\infty)$, and $(F_n - \mathbb{E}F_n)/\sqrt{\operatorname{Var}(F_n)}$ converges in law to a nondegenerate probability measure with mean zero and variance one. The proofs in [82] use the exponent of Theorem 1.6 and the comparison estimates of [81]; see [82, Appendix A]. We do not use [82] in this paper. On the side of large deviations, Pei identified the coefficient of the fifth-order term in the constrained Parisi functional and the resulting rate for upward deviations of F_n/n, with $n \to \infty$ taken before the deviation tends to zero [95, Theorem 1].

The proof of Theorem 1.6 in [81] reduces it to a two-sided bound on a positive logarithmic moment. Let $F_n^c = F_n + \beta G$, where G is a centered Gaussian variable of variance $1/2$, independent of the disorder. The variable βG completes the covariance of βH_n to $n\beta^2 R(x,y)^2/2$, and we call F_n^c the completed log partition function (Section 3.2). For every fixed $m \in (0, 1/300)$ and $s_n = n^{-1/6+5m}$, it is shown in [81, Proposition 7.5] that $\log \mathbb{E}\exp\{s_n(F_n^c - \mathbb{E}F_n^c)\}$ lies between constant multiples of n^{20m} and n^{100m} for all sufficiently large n. We state this bound precisely as Proposition 20.1. The deduction of Theorem 1.6 from this bound in [81, Lemmas 7.6 and 7.7] uses an argument of Aronow and Lopatto [8, Proposition 5.1], based on the log-concavity of an exponential augmentation of F_n^c. We reproduce it in Section 20.

The mean bias. For fixed $\beta > 1$, Aronow and Lopatto proved that $c_\beta n^{-11/12} \le \Delta_{n,\beta} \le C_\beta n^{-2/3}$ for all sufficiently large n [8, Theorem 1.4], and later that $\Delta_{n,\beta} \le C_b n^{-7/10}$ for every $n \ge 2$ [6, Theorem 1.1]. Their lower bound $n^{-11/12}$ follows from (1.7) and the lower bound $c_\beta n^{4/15}$ on the variance; with (1.5) in place of this bound, the exponent becomes $5/6$, as in Corollary 1.7. The proof of (1.7) balances a gain that is linear in a small positive tilt and proportional to the variance against the fifth-order cost of constraining the Parisi measure to have mass at least s at zero [8, Lemma 7.2]. In physics, the correction is expected to be of order n^{-a} with $a < 1$ at low temperature, and two values have been proposed. The value $a = 2/3$ was supported by the zero-temperature numerics of Bouchaud, Krzakala, and Martin [27] and of Boettcher [22], by Billoire's numerical integration along the interpolation of Guerra and Toninelli [60] in [17], and by the argument of Aspelmeier, Billoire, Marinari, and Moore mentioned above. Franz, Parisi, and Ricci-Tersenghi argued that the correct value is $a = 3/4$, possibly up to logarithmic factors [56, equations (13)–(17)]. The bound of [6] excludes a nonzero correction of order $n^{-2/3}$ at every fixed $\beta > 1$ and is consistent with the value $3/4$. A fuller account of these predictions is given in [6, Section 1.4].

Normalizations. We have converted the results quoted in this section and the previous one to the normalization (1.1). Some of the cited works sum over ordered pairs of sites, which changes the inverse temperature by a constant factor; for instance, the critical inverse temperature of [38] is $1/\sqrt{2}$ in its own normalization. Section 3.6 compares our normalization with those of the works on which our proofs rely.

1.5 Formal verification

OpenAI has published formalizations of [77] and [81] in the Lean 4 proof assistant [48, 83, 108]. The main formal theorem on the fluctuation exponents takes the two-sided bound of Proposition 20.1 as an explicit hypothesis. In [81], that bound is the conclusion of Proposition 7.5, which is proved there on paper. We have completed this formal proof: our Lean development proves the three conclusions of Theorem 1.6 with no hypothesis beyond the definition of the model and $\beta > 1$. It follows the mathematical route of [81]; the work consisted in carrying out that route in full detail for the comparison problems of [81, Section 3] whose minimizers have a positive lower radius, and for the successive comparisons built from them.

We have also formalized Theorem 1.1, Corollary 1.2, and Corollary 1.7. In these formal proofs, the upper bound on the mean bias of [6, Theorem 1.1] and the variance–bias inequality (1.7) are proved, not assumed; their formal proofs come from our Lean development [7] for [6]. The formal proof of Corollary 1.7 also proves the identities that relate the normalizations of Theorem 1.6 and (1.7). The formal proof of Theorem 1.1 has the structure of the written proof, but it uses a weaker form of the local expansion of Proposition 7.13, which suffices for the same exponent. Theorem 12.6 is not formalized.

In their source packages, each of these four formal theorems was compared with a separately written statement of the same result that depends only on Mathlib, and its proof was checked to use only the standard axioms `propext`, `Classical.choice`, and `Quot.sound`. The declarations on which its statement and proof depend were then checked again by three kernels: the kernel of Lean and the independent kernels nanoda and con-ron. Section 22 states the formal theorems, explains how they correspond to the statements of this paper, describes what the completion of the proof of Theorem 1.6 had to establish, and describes these checks.

Theorem 1.3 is also formalized, as the statement described in Section 22.5, and is included in the combined five-claim verification described in Section 22. Corollaries 1.4 and 1.5 are not separately formalized. The proofs in this paper are complete without reference to the formalization.

1.6 Methods and their sources

We describe the proofs of Theorems 1.1 and 1.3 and the sources of their ideas. A more detailed overview is given in Section 2.

The lower bound

From overlap constraints to slow mixing. We follow the strategy of [77]. Its first step reduces the problem to a statement about the stationary chain [77, Section 8.1]. Let v be a reference configuration with $R(v, X_0)$ bounded away from zero. By spin-flip symmetry, the half-space $\{y : R(v, y) \le 0\}$ has Gibbs mass at least $1/2$. If $d^J(X_0, t) \le \tau < 1/2$, the chain started at X_0 is in this half-space at time t with probability at least $1/2 - \tau$. A single update changes $R(v, \cdot)$ by at most $2/n$, so the chain must cross every intermediate overlap level. It suffices to show that the chain started from $X_0 \sim \pi^J$ crosses these levels before time t only with small probability. At each deterministic time the stationary chain has law π^J, and a union bound over times controls the discrete chain; an independent Poisson clock of rate n then gives the continuous-time chain. This argument is related to bottleneck estimates for Markov chains (see [68, Theorem 7.4]). Here, however, the bottleneck depends on the initial configuration, and it must separate almost every initial configuration in Gibbs measure from equilibrium.

To show that crossings are rare, we build, as in [77], a growing family of reference configurations along the path. The references are chosen from independent Gibbs samples, and their overlaps with the current configuration are constrained simultaneously. At each crossing level, we find a new reference with large overlap with the current configuration using a bound on the Gibbs mass of overlap neighborhoods [77, Proposition 6.1]. This bound is derived from an upper tail bound for the energy and an estimate for small Gaussian rotations of the disorder [77, Lemmas 5.1 and 5.2]. We then search backward for the crossing of the overlap level of the new reference, while overlap constraints keep the overlaps with the earlier references almost unchanged. Iterating this step creates many simultaneous constraints at a single time. We arrange the references in blocks of fixed length, so that a conflict deletes only a short final segment of a block, and we draw the overlap thresholds at random, one level at a time; these features allow us to obtain the final probability estimate [77, Sections 8.3–8.5].

The overlap constraints have the following form. Let ε be a tolerance. If the mean r of the overlaps $R(v, x)$ and $R(v, w)$ lies in a fixed compact subinterval of $(0, q_\beta)$ and $R(w, x) \ge r + \rho$, then $|R(v, x) - R(v, w)|$ does not lie in $[2\varepsilon, 4\varepsilon]$. In words, a large overlap between w and x forces their overlaps with v to agree. Since overlaps change by at most $2/n$ in one update, excluding this band keeps the two overlaps close along a path. To show that such a constraint fails with small Gibbs probability, we show that the partition function restricted to the triples (v, x, w) with violating overlaps is small compared with

Z_n^3. In physics, free energies constrained by the overlap with an equilibrium reference configuration go back to Franz and Parisi [55]. The work of Ben Arous and Jagannath [16] is an earlier rigorous argument that passes from such constrained free energies to the dynamics, there for the spectral gap.

Constrained comparisons and the scalar problem. We bound constrained partition functions by Gaussian interpolation. Guerra's hierarchical interpolation [59] gives the upper bound in the Parisi formula. Talagrand extended it to systems of replicas with prescribed overlaps, in the proof of the matching lower bound [103, Section 3] and in the two- and three-replica bounds of [105, Propositions 3.4 and 4.4]. These bounds already retain a cost when the prescribed overlaps differ from the endpoint of the interpolation. Bovier and Klimovsky developed the interpolation for multidimensional spins [29]. Chen studied the two-dimensional Guerra–Talagrand bound and its variational representation [36], and Panchenko proved the vector-spin variational formula and the asymptotic sharpness of the constrained bounds [89]. We use the matrix comparison for finite hierarchies of [77, Proposition 2.3] for the quadratic covariance of SK, which allows any finite number of replicas and endpoint matrices whose diagonal entries need not equal one. The hierarchies themselves descend from Parisi's recursion [91] and from Ruelle's probability cascades [97]; see [88]. For the mixed even p-spin models with exponentially summable coefficients, including SK, Aronow and Lopatto identified the limiting overlap array averaged over the disorder [5, Theorems 1.1 and 1.2]. At zero field, the absolute overlaps have the law of the overlap array $(U_{ab})_{a<b}$ of the corresponding Ruelle probability cascade, and the signed array has the law of $(\epsilon_a\epsilon_b U_{ab})_{a<b}$, where the replica signs ϵ_a are independent symmetric signs, independent of the cascade. These are limiting statements, while the proof below needs estimates at finite n on shrinking scales.

We obtain the pressure deficits from the scalar Parisi problem. Auffinger and Chen proved that the Parisi functional has a unique minimizer, and they expressed it through a stochastic control problem [13, Theorems 2 and 3 and Corollary 1]. They also proved regularity properties of the Parisi equation and of its minimizers [12, Propositions 1 and 2 and Theorems 1 and 2]. Jagannath and Tobasco gave a dynamic-programming treatment of the functional [62, Lemma 18] and characterized its minimizers through their first variation [63, Proposition 1.1 and Corollary 3.10]; Chen obtained the same characterization [36, Theorem 2 and Proposition 1]. Toninelli proved that replica symmetry is broken at zero field for $\beta > 1$ [111]. Let u be the spatial derivative of the solution of the Parisi equation at the minimizer, and let Y be the associated diffusion. The identity $\mathbb{E}u(t, Y_t)^2 = t$ at every point t of the support is due to Talagrand [104]; see also [12, Theorem 5], [36, Proposition 1], and [69, Lemma 2.2]. OpenAI proved it in the normalization used here, together with the regularity of the scalar evolution that it requires [77, Lemma 3.1 and Proposition 3.2]; the argument of [77] does not assume a description of the support, and it treats separately the case of a forbidden overlap interval [77, Section 8.2]. By [69,

Theorem 1.1], the identity holds on the whole interval $(0, q_\beta)$, and this case does not arise. Differentiating the identity once and twice along this interval, we obtain two stationarity identities,

$$\beta^2 \mathbb{E}\chi(t, Y_t)^2 = 1, \qquad \mathbb{E}w(t, Y_t)^2 = 2\alpha(t)\mathbb{E}\chi(t, Y_t)^3 \qquad (0 < t < q_\beta),$$

where χ and w are the second and third spatial derivatives and α is the distribution function of the Parisi measure (Section 5). Both identities appear, in the normalization of energy units, in the proof of [8, Lemma 7.2] (see [8, equation (7.12)] and the identity preceding it), and the differentiated moment identity is also in [69, Lemma 2.5]. The first identity expresses, in scalar form, the marginal stability of the Parisi solution, which De Dominicis and Kondor found in a truncated replica model [47]; Crisanti and Rizzo wrote it in physical variables [44, equations (22) and (23)].

Positive fractional moments $\mathbb{E}Z_n^s$ have their own genealogy. Talagrand computed the limits of their normalized logarithms [106]. For SK at zero field and fixed $s \in (0, 1)$, the limit of $(ns)^{-1} \log \mathbb{E}Z_n^s$ is the minimum of the Parisi functional over measures with mass at least s at zero. This formula is contained in the multiscale formula of Contucci and Mingione [43, Theorem 2.1], and it was stated in this form by Chen, Guionnet, Ko, Lacroix-A-Chez-Toine, and Mourrat [35, Theorem 2.1]. Aronow and Lopatto made the comparison quantitative when s depends on n [8, Theorem 1.1], and they proved that the constraint raises the Parisi value by an amount of order s^5 [8, Lemma 7.2 and Proposition 7.3]. This is a rigorous counterpart of the replica calculations of Kondor and of Parisi and Rizzo, and Pei identified its leading coefficient [95, Theorem 1(a)].

What is new. Four parts of the proof are new. Together they allow overlap constraints at the scale $\rho \asymp n^{-7/30}$, with tolerance comparable to ρ.

1. *A three-replica covariance perturbation with a cubic deficit* (Section 7). In [77, Section 7], the overlap constraints near a level r come from a local perturbation of the covariance path of width ρ^8 with tolerance ρ^{20}, and the resulting pressure deficit is of order ρ^{43}. We keep the block hierarchy and the antisymmetric Gaussian contrast of [77, Section 7, Steps 1 and 2], but we perturb the covariance path differently. It is positive semidefinite at every stage, ends exactly at the prescribed overlap matrix after a terminal correction, and its expansion uses both stationarity identities to cancel the lower-order terms. The resulting deficit is of order ρ^3 per spin, with tolerance comparable to ρ, uniformly for r in a compact subinterval of $(0, q_\beta)$. Its coefficient involves $\mathbb{E}\chi(r, Y_r)^3$, which is positive (Section 5).
2. *A positive fractional-moment comparison* (Section 8). For constraints near the level r, the deficit $n\rho^3$ in the logarithm of a constrained partition function is smaller than the scale $O_\beta(\sqrt{n})$ supplied by the standard Gaussian concentration bound for Lipschitz functions, so that bound is too coarse here. Instead, we bound the moment

$\mathbb{E}Z^s$ of a constrained partition function Z, with s comparable to ρ, by a fractional-moment version of the matrix comparison of [77, Proposition 2.3], in which the outer level of the hierarchy has mass s, as in the positive-moment interpolation of [8, Section 2]. The comparison uses the profile $\max\{\alpha, 3s\}$, whose cost is of order s^5 by the trial bound of [8, Lemma 7.2], together with a bound on the response of the local perturbation to this change of profile. Markov's inequality then bounds the disorder probability of failure of each of these bounds on constrained partition functions by $e^{-cn\rho^4}$.

3. *One disorder event for all constraints* (Section 9). The denominator Z_n^3 is bounded below on the event $F_n/n \geq P_\beta - c_0\rho^3$, for a fixed $c_0 > 0$. Let $\delta \in (0, c_0)$ be fixed such that $\Delta_{n,\beta} \leq (c_0 - \delta)\rho^3$ for all sufficiently large n; this margin is obtained by choosing the constant R in $\rho = Rn^{-7/30}$ large. Chebyshev's inequality and the variance bound of [8, Theorem 1.4] give

$$\mathbb{P}\{F_n/n < P_\beta - c_0\rho^3\} \leq \frac{\mathrm{Var}(F_n)}{\delta^2 n^2 \rho^6} \leq C_{\beta,\delta}\, n^{-23/15}\rho^{-6},$$

which is $O(n^{-2/15})$ for $\rho = Rn^{-7/30}$. We intersect this event with the bounds on polynomially many constrained partition functions, and the complement of the resulting event, whose probability tends to zero, is discarded once and for all. On this event, the number of reference configurations and path steps, which is exponential in $1/\rho$, multiplies only the conditional Gibbs probability of a violation, which is at most $e^{-cn\rho^3}$. This replaces the probability estimate of [77, Lemma 4.2].

4. *Control of crossing errors through the next retained reference* (Sections 11 and 12). In [77, Section 8.4], the overlap of an earlier reference with the current configuration may drift by up to $6j\varepsilon$ after j levels, and the final count requires $p\varepsilon/\rho$ to be small for p levels. Since p is of order $1/\rho$, this forces ε to be much smaller than ρ^2. Instead, we control the overlap of each retained reference with the next retained reference. This overlap changes only when a deletion occurs, and the error stays of order ε at every level, rather than accumulating. The first reference of each block keeps its next retained reference, and at one common time the overlap of the path with the first reference of each of the M blocks lies within a fixed multiple of ε of its threshold. A count over the $m+1$ path steps of these M windows of predictable center gives the bound $(m+1)\gamma_n^M$, where γ_n is the ratio of the window length to the length of a threshold interval, and γ_n stays below one when ε is a small fixed multiple of ρ. This allows ε comparable to ρ.

The other overlap constraints use the inherited pressure bounds: the pair bound of [77, Lemma 4.3], the three-replica bound at fixed scales of [77, equation (4.20)], and, for large overlaps $R(w, x)$, the comparison of [77, Section 7, Steps 5–8]. We also extend the bound

on the Gibbs mass of overlap neighborhoods of [77, Proposition 6.1] to every scale $L \to \infty$ with $L = o(n)$; its proof is that of [77] (Section 10).

The exponent. We determine the exponent by balancing these estimates. The lower bound on Z_n^3 in item 3 requires the mean bias $\Delta_{n,\beta}$ to be smaller than a small multiple of the deficit ρ^3. The comparison proved in [77, Theorem 2.4] has error $O(n^{-1/24})$. This is much smaller than the deficit ρ^{43} at the scale $\rho = n^{-1/10000}$ used there, but not than ρ^3 at the scale used here. For this reason we use the upper bound $\Delta_{n,\beta} \le C_b n^{-7/10}$ of Aronow and Lopatto [6], and we take $\rho = Rn^{-7/30}$ for a large constant R. A fixed overlap interval inside $(0, q_\beta)$ then contains of order $1/\rho \asymp n^{7/30}$ crossing levels. By counting, we find that the probability of crossing all of them before time t is small when $\log t$ is a small multiple of the number of levels. This leads to the time $\exp(A_\beta n^{7/30})$. Two further conditions concern the two kinds of union bound in the proof. For the union over reference configurations and path steps, the conditional Gibbs error is at most $\exp(C/\rho - cn\rho^3)$, which tends to zero if $cn\rho^3 - C/\rho \to \infty$. For the disorder event, a union over at most Cn^d constrained partition functions has failure probability at most $Cn^d e^{-c'n\rho^4}$, which tends to zero if $c'n\rho^4 - d\log n \to \infty$. The condition $n\rho^4/\log n \to \infty$ implies both. For $\rho = Rn^{-\lambda}$ with fixed R, it holds when $\lambda < 1/4$, in particular at $\lambda = 7/30$; at $\lambda = 1/4$, the second union bound fails. This is the restriction to exponents below $1/4$ mentioned in Section 1.1.

Inputs for the lower bound. Besides standard tools, the proof of Theorem 1.1 uses the following public results, which are stated in Section 4 with their hypotheses: the Parisi formula and Guerra's bound [59, 103], the uniqueness of the Parisi minimizer [13], the interval support of the Parisi measure [69], the upper bound on the mean bias of [6, Theorem 1.1], the upper bound on the variance and the trial bound of [8], and results of [77, Sections 2–8]. The proof of Theorem 12.6 uses in addition the small-deviation inequality of [90]. All other steps are proved in this paper.

The upper bound

From flows to mixing. We follow the strategy of [85]. Its last step reduces the problem to a lower bound on stationary flows across cuts [85, Section 7, Lemma 8]. For $S \subseteq \Sigma_n$, the flow $\sum_{x\in S,\, y\notin S} \pi^J(x)\mathcal{L}^J(x,y)$ is the stationary rate at which the continuous heat-bath process jumps from S to its complement. Let $0 < \epsilon < 1/3$ and $\phi > 0$, and suppose that the flow across S is greater than ϕ for every S with $\epsilon \le \pi^J(S) \le 1 - \epsilon$. Then for every $t > 0$,

$$\sum_{x\in\Sigma_n} \pi^J(x)\, d^J(x,t) \le \frac{n\sqrt{2\log 2}}{\phi\sqrt{t}} + 2\epsilon.$$

The proof removes a set of small Gibbs mass, proves a Poincaré inequality on the remainder by the coarea argument for the reversible Cheeger bound [67, Theorem 3.5], and

then bounds the Gibbs average of the distances through the entropy dissipation of the whole chain [51, Lemmas 2.5 and 2.7]. It uses only cuts that separate appreciable Gibbs masses, and it does not require the trajectory to stay in the retained set. We reproduce it in Lemma 19.1. With $\epsilon = n^{-\mathsf{d}}$, $\phi = e^{-n^{1-\mathsf{q}}}$, and $t = e^{n^{1-\omega}}$, where d and q are positive exponents of the proof (Section 13.3), the right side tends to zero when $\omega < \mathsf{q}$, and using Markov's inequality for the Gibbs average, we obtain Theorem 1.3. It therefore suffices to show that, with probability tending to one over the disorder, the flow across every S with $\min\{\pi^J(S), \pi^J(S^{\mathrm{c}})\} \geq n^{-\mathsf{d}}$ is greater than $e^{-n^{1-\mathsf{q}}}$.

The proposal. To bound these flows, OpenAI constructs a random configuration correlated with the disorder, called a proposal [85, Section 4]. Each coordinate of the proposal follows the magnetization martingale of the scalar Parisi diffusion. The martingale is driven by successive queries of the interaction matrix: over each cell of a time mesh, the increment of the magnetizations, orthogonalized against the earlier query directions, is multiplied by the matrix. By orthogonal invariance, the part of the matrix that has not been revealed is a compressed block of the Gaussian orthogonal ensemble (GOE) at each step. Beyond the top q_β of the support of the Parisi measure, the martingale is continued until the average uncertainty of the coordinates is small, and the query noise is mixed with an independent Brownian noise. At the final time, the spins are drawn independently with the magnetizations as means. Successive Gaussian conditioning along orthogonalized directions is used in Bolthausen's iterative construction of solutions of the TAP equations [25, Sections 2–3], and Montanari's incremental approximate message passing algorithm uses a Parisi diffusion and normalized increments of the magnetizations [75, Sections 2–3].

The equilibrium input is the stationarity of the Parisi measure along the whole interval $[0, q_\beta]$. The stationarity conditions on the support were obtained by Chen [36, Theorem 2 and Proposition 1] and by Jagannath and Tobasco [63, Proposition 1.1 and Corollary 3.10], and the support is the whole interval by [116, Theorem 1] for β close to one and by [69, Theorem 1.1] for every $\beta > 1$. Along the scalar diffusion, an exact identity balances the information about the matrix that the queries reveal against the energy gained by the proposal [85, Section 3, Lemma 4]. It yields the two estimates of [85, Proposition 5]. The relative entropy of the joint law of the matrix and the proposal with respect to the joint law of the disorder and a Gibbs sample is sublinear in n, and outside an event of small probability, the law of the pair is dominated by $e^{o(n)}$ times this joint law.

Coverage and the connecting sequence. The information inequalities used in these comparisons—the chain rule, data processing, and the variational formula for relative entropy—are recalled in [28, Chapter 4]. Forward relative entropy alone does not show that the proposal reaches a prescribed set of positive Gibbs mass. Under the joint law of the disorder and a Gibbs sample, the conditional law of the disorder given the spin is

strongly log-concave. A transport bound for such laws [24, Corollary 3.10] and Gaussian smoothing of the disorder [70, Lemma 11] reverse the entropy comparison. They give a lower bound on the probability that the proposal lands in any set of appreciable Gibbs mass [85, Section 5, Proposition 6]. Two proposals generated from a common smoothed disorder then land on the two sides of a balanced cut with a probability that is not too small. Following [85], these two proposals are run to a longer, fine horizon, and they are joined by a sequence of proposals run to a shorter, coarse horizon, obtained by rotating the Gaussian inputs, with small consecutive Hamming distances. Near-pinning of the magnetizations joins the two horizons. If the cut had small flow, its enlarged vertex boundary would have tiny Gibbs mass, and the likelihood bound keeps the coarse proposals away from this boundary. But a sequence with short steps that crosses the cut must come close to its boundary. This contradiction bounds the flows from below [85, Section 6, Proposition 7].

What is new. In [85], the parameters of the proof are fixed numbers, and the saving $1/40000000$ is not optimized [85, Section 1, after Theorem 1]. To reach every saving below $1/4$, the estimates for the coarse proposal must hold outside events of probability close to $e^{-n^{3/4}}$, with a likelihood bound close to $e^{n^{3/4}}$ times the joint law. The following refinements yield every fixed saving below $1/4$.

1. *Sharp estimates for the coarse proposal* (Sections 16 and 17). We strengthen the scalar entropy–energy identity of [85, Section 3, Lemma 4] to a pathwise terminal identity. Along the scalar diffusion beyond the support, the information revealed by the queries and β times the linearized energy of the proposal cancel up to a terminal term. The centered fluctuation of this term has variance at most $\beta^4 p(V)^3$, where $p(V)$ is the mean conditional variance of a spin given its magnetization at the final time V (Lemma 17.5). The coarse likelihood estimate absorbs the deterministic amplitude of this term. The centered difference between this cancellation and its version on the mesh has a moment generating function bounded by $e^{C_\beta \theta^2 h}$ for $|\theta| \le c_\beta / h$, where h is the mesh parameter (Lemma 17.6). We combine these bounds with concentration of weighted dyadic sums over cells of unequal length, normalized by the durations of the cells (Lemma 16.6), with a bound on the deviation of the Gram matrix of the queries from the identity that uses the variance of the squared projections and not only their tails (Lemma 16.1), and with sharper bounds on the variances of the Gaussian terms that the proposal does not use (Lemma 17.2). The dyadic triangular and Cholesky arguments build on [85, Section 4, proof of Proposition 5].
2. *The growing mesh* (Section 14). The cells of the mesh in [85] have a common length. Ours have length $he^{t/4}$ at time t, with $h = n^{-\mathsf{a}}$, so that the horizon $L \log n$ is reached after at most $8n^{\mathsf{a}}$ cells when $\mathsf{a} > L/4$ (Lemma 14.3). The quadrature error is $O_\beta(h)$, uniformly in the horizon. On this mesh, adjacent cells have different lengths, and the

estimates distinguish the duration of the previous cell, which governs the conditional law of the spins, from that of the next cell, which normalizes the query. For instance, the bound on the conditional correlations of the query entries with the spins in Section 15.2 loses a factor two compared with equal cells.

3. *Direct expected energy at the fine horizon* (Section 15.2). We compute the expected queried energy through the exact one-query conditional identity of Lemma 15.3, including fallback outcomes. This avoids the Gram-event restriction in the energy calculation of [85, Section 4, Proposition 5] and gives the fine relative-entropy bound of Proposition 15.5, with saving $\mathsf{k_f} < \min(\mathsf{a}, \mathsf{f}, 1-\mathsf{a})$.
4. *The parameter system* (Section 19.5). The exponents of the proof are affine functions of $\varepsilon_\omega = (1/4 - \omega)/20$, and this single family satisfies every condition of the proof for every $\omega \in (0, 1/4)$ (Proposition 19.2).

The connecting sequence, including fine proposals only at the two endpoints, stability from coarse Gram events, and pinning without a fine Gram exception, follows [85, Section 6, proof of Proposition 7].

The exponent. The proof of Proposition 19.2 shows that, for every choice of the exponents, the estimates of Part IV require

$$\omega < \mathsf{q} < \mathsf{s_b} < \mathsf{s_{seq}} < \mathsf{e} < \mathsf{u_E} < \mathsf{k_f}/2 \le 1/4.$$

The mixing step needs $\omega < \mathsf{q}$, since the flow threshold is $e^{-n^{1-\mathsf{q}}}$. In the connecting sequence, $e^{-n^{1-\mathsf{s_b}}}$ bounds the failure probability of one coarse proposal, $e^{-n^{1-\mathsf{s_{seq}}}}$ that of the whole sequence, and $e^{-2n^{1-\mathsf{e}}}$ bounds from below the probability that the two fine endpoints lie on opposite sides of the cut, and the proof needs $\mathsf{q} < \mathsf{s_b} < \mathsf{s_{seq}} < \mathsf{e}$. Smoothing and coverage need $\mathsf{e} + \mathsf{d} < \mathsf{u_E} < \min\{\mathsf{D}, \mathsf{k_f} - \mathsf{D}\}$, where $n^{-\mathsf{D}}$ is the smoothing parameter and $n^{1-\mathsf{k_f}}$ bounds the forward relative entropy at the fine horizon. Finally, this relative entropy contains terms of order $nh = n^{1-\mathsf{a}}$ and n^{a}, the second from the number of queries, so that $\mathsf{k_f} < \min\{\mathsf{a}, 1-\mathsf{a}\} \le 1/2$.

Inputs for the upper bound. Besides standard tools of Gaussian analysis and stochastic calculus [28, 64, 112], the proof of Theorem 1.3 uses the following results, which are stated in Section 4 and Section 13.4 with their hypotheses: the Parisi formula and Guerra's bound (Theorem 4.4), the uniqueness of the Parisi minimizer (Theorem 4.5), the interval support of the Parisi measure (Theorem 4.7) and the identity $\Gamma(t) = t$ on $(0, q_\beta)$ (Corollary 4.8), the regularity and scalar evolution of Proposition 4.3, the scalar estimates of [85, Section 3, Lemmas 3 and 4], and the transport, smoothing, and finite-chain results cited above. All other steps are proved in this paper.

The fluctuation theorem

The proof of Proposition 20.1 in [81] combines finite replica trees and covariance differentiation for Gaussian hierarchies [81, Section 2], optimized short comparisons over overlap quantiles [81, Section 3], a one-site cavity estimate, which compares an overlap moment of a comparison with the corresponding moment of a scalar recursion [81, Proposition 4.1], a rigidity theorem, by which rescaled overlap quantiles of minimizing comparisons have linear limits at small masses [81, Theorem 5.1], finite propagation of regularity through successive comparisons [81, Proposition 6.6], and a low-field perturbation that yields the upper bound [81, Section 7]. As explained in [81, Section 1], its closest analytic predecessor is the combination in [8] of a pinned Brascamp–Lieb estimate for the variance [30], optimized covariance interpolation, and positive moments [8, Propositions 2.2 and 2.3 and Section 3]. The single-site analysis of Thouless, Anderson, and Palmer [109] is an early antecedent of the cavity method in spin glasses, and Mézard, Parisi, and Virasoro developed the cavity construction of the solution with broken replica symmetry [74]. The cavity variational principle of Aizenman, Sims, and Starr [2] is a rigorous antecedent of comparisons that add or remove sites. Section 20 summarizes this route, with references to the corresponding statements.

1.7 Organization of the paper

Section 2 gives an overview of the proofs of both bounds. Part I contains preliminaries: Section 3 defines the model, the dynamics, and the joint law of the disorder and the initial configuration, and Section 4 defines Gaussian hierarchies and the Parisi functional and states the imported results. Part II contains the static estimates for the proof of Theorem 1.1. Section 5 treats the scalar problem near a point of the support of the Parisi measure, Section 6 proves short-time estimates for Gaussian hierarchies, Section 7 constructs the three-replica covariance perturbation and proves the cubic deficit, Section 8 proves the fractional-moment comparison, and Section 9 combines these into a single estimate for all overlap constraints. Part III contains the dynamical part of the proof of the lower bound. Section 10 treats the Gibbs mass of overlap neighborhoods and the families of reference configurations, Section 11 proves that the overlap constraints persist along a trajectory, and Section 12 completes the proof of Theorem 1.1 and Corollary 1.2 and proves Theorem 12.6.

Part IV proves the upper bound. Section 13 introduces the Gaussian matrix from which the disorder is built, the scalar Parisi process and its continuation beyond the support, and the exponents of the proof. Section 14 constructs the Gaussian-query proposal on a growing mesh, and Section 15 bounds its forward relative entropy at the fine horizon. Section 16 proves the concentration estimates on the growing mesh, and Section 17 proves the joint cancellation and the likelihood bound for the coarse proposal.

Section 18 proves the coverage estimate and constructs the connecting sequence, and Section 19 proves Theorem 1.3 and Corollaries 1.4 and 1.5 and verifies the conditions on the exponents.

Part V concerns the fluctuations and the mean bias: Section 20 states the bound of Proposition 20.1 and deduces Theorem 1.6 from it, and Section 21 proves Corollary 1.7. Part VI contains Section 22, which describes the formal proofs, and Section 23, which lists open problems.

SECTION 2

Overview of the proofs

This section gives an overview of the proofs of Theorem 1.1 and Theorem 1.3, and a shorter account of the parts of the paper that concern Theorem 1.6 and Corollary 1.7. Every claim made here is proved later in the paper, or quoted there with its source. Sections 2.1 to 2.3 explain the reduction to paths, the architecture of the proof in [77], which we follow, and the balance of scales that gives the exponent $7/30$. Sections 2.4 to 2.6 describe the cubic pressure deficit (Sections 5 to 7), the passage to Gibbs probabilities (Sections 8 and 9), and the dynamical argument (Sections 10 to 12). Section 2.7 concerns Part IV, and Section 2.8 concerns Part V. Table 2.1 shows how the sections depend on each other.

Throughout this section, $\beta > 1$ is fixed. In Sections 2.1 to 2.6, $\pi = \pi^J$ is the Gibbs measure at a fixed disorder. We write λ for the exponent of the barrier, and we use an upper bound $\Delta_{n,\beta} \leq C_b n^{-3\lambda}$ on the mean bias $\Delta_{n,\beta} = P_\beta - n^{-1}\mathbb{E}F_n$. For Theorem 1.1, we take $\lambda = 7/30$ and the bound $\Delta_{n,\beta} \leq C_b n^{-7/10}$ of Theorem 4.11. The exponent of the upper bound is $1 - \omega$, for a fixed $\omega \in (0, 1/4)$, and Section 2.7 uses the notation of Part IV.

2.1 The sign test

Recall from Lemma 3.4 that Theorem 1.1 states that the Gibbs mass of the initial configurations from which the chain is at total variation distance greater than τ from π at time $e^{A_\beta n^{7/30}}$ tends to one in probability over the disorder. Since π is stationary, the chain started from a Gibbs sample has law π at every deterministic time (Lemma 3.2), and the law averaged over the initial configuration carries no information. A lower bound on the distance must come from a property of almost every individual initial configuration that the dynamics forgets slowly.

The property we use is the sign of an overlap. As in [77, Section 8], we choose for each $x \in \Sigma_n$ a reference configuration $v(x)$, with $R(v(x), x) \geq q_*$ for all x outside a set of small Gibbs mass, where $q_* \in (0, 1)$ is a fixed overlap (Definition 11.2). Since $H_n(-y) = H_n(y)$, the half-space $\{y : R(v(x), y) \leq 0\}$ has Gibbs mass at least $1/2$. If the law ν_x of the chain started at x is within τ of π in total variation, then the half-space has mass at least

Table 2.1. The main results of Sections 3 to 21, the sections whose results they use, and the sections in which they are used.

Section	Main results	Uses	Used in
3 Model	realized-start probability, Lemma 3.4; heat-bath dynamics, Lemma 3.2; completed Hamiltonian, Lemma 3.1	—	4, 8, 10, 12, 13, 18, 19, 20, 21
4 Inputs	matrix comparison, Proposition 4.19; upper bound on the mean bias, Theorem 4.11; variance and profile bounds, Theorem 4.10 and Proposition 4.12; pair, fixed-scale, and separated-overlap bounds, Propositions 4.23 to 4.25; overlap-neighborhood masses, Proposition 4.29	3, literature	5–10, 12, 13, 17, 18
5 Scalar problem	stationarity identities, Lemma 5.3; the constant m_β, Corollary 5.5	4	7, 8, 10, 13, 17
6 Flows	uniform short-time bounds, Proposition 6.3; response to the terminal function, Lemma 6.13 and Corollary 6.14	4	7
7 Cubic deficit	local expansion, Proposition 7.13; deficit on a band, Proposition 7.15	4–6	8, 10
8 Fractional moments	fractional comparison, Proposition 8.3; local fractional bound, Lemma 8.7	3–5, 7	9
9 Static event	simultaneous disorder event, Lemma 9.7	4, 8	10, 12
10 References	overlap-neighborhood masses, Lemma 10.1; fixed families of references, Lemma 10.4	3–5, 7–9	11, 12
11 Induction	constraints along a path, Proposition 11.10; a common time, Corollary 11.11	10	12
12 Barrier proof	sign obstruction, Proposition 12.2; Theorems 1.1 and 12.6 and Corollary 1.2	3, 4, 9–11	19
13 Upper-bound setup	primitive disorder and Gaussian bounds, Lemmas 13.2 and 13.3; the scalar process beyond the support, Proposition 13.4 and Lemmas 13.5 and 13.6	3, 4, 5	14–19
14 Proposal	Gaussian assembly, Lemma 14.1; spin laws, Definition 14.2; growing mesh, Lemma 14.3	13	15–18
15 Fine entropy	reference law, Lemma 15.1; forward entropy, Lemma 15.2 and Proposition 15.5	13, 14	17, 18
16 Concentration	fixed Gram stability, Lemma 16.1; weighted concentration, Lemma 16.6	13, 14	17
17 Coarse domination	weighted energy, Lemma 17.2; joint cancellation, Lemmas 17.5 and 17.6; likelihood bound, Proposition 17.8	4, 5, 13–16	18
18 Connections	smoothing and coverage, Lemmas 18.2 and 18.3; boundary mass, Lemma 18.4; connecting sequence, Proposition 18.8; witness decay, Proposition 18.9	3, 4, 13–15, 17	19
19 Upper-bound proof	pruning, Lemma 19.1; Theorem 1.3 and Corollaries 1.4 and 1.5; parameters, Proposition 19.2	3, 12, 13, 18	—
20 Fluctuations	the bound of [81], Proposition 20.1; log-concave augmentation, Lemma 20.2; Theorem 1.6	3	21
21 Mean bias	variance–bias inequality, Proposition 21.1; Corollary 1.7	3, 20	—

$1/2 - \tau$ under ν_x. Multiplying by $\pi(x)$ and summing over x, we find that

$$\left(\frac{1}{2} - \tau\right) \pi\{x : d^J_{\mathrm{disc}}(x,k) \le \tau\} \le \mathbb{P}\{R(v(X_0), X_k) \le 0\}, \tag{2.1}$$

where $(X_k)_{k\ge 0}$ is the discrete chain with $X_0 \sim \pi$ (see (12.8)). This is the reduction of [77, (8.1)–(8.3)]. The half-space depends on the initial configuration, so that the distance on the left side is evaluated before the average over X_0, while the right side is a probability for the stationary chain. Since one update changes every overlap by at most $2/n$, a chain that enters the half-space must cross every overlap level between q_* and 0.

We show that these crossings cannot occur along paths that avoid a fixed set of small Gibbs mass. We call a sequence $x_0, \dots, x_m$ a path if all overlaps change by at most $2/n$ from one element to the next (11.2), so that every trajectory of the discrete chain is a path. Let $L = An^\lambda$ be the logarithmic time scale, for a coefficient $A > 0$ fixed below, so that e^L is the time horizon, and let $T_n = \lceil n^2 e^L \rceil$ be the budget of update attempts. For every disorder in an event $\mathcal{G}_n$ with $\mathbb{P}(\mathcal{G}_n) \to 1$, we construct the map $x \mapsto v(x)$ and a set $V \subseteq \Sigma_n$ with $\pi(V) \le 3e^{-4L}$ such that

$$R\big(v(x_0), x_m\big) > 0 \qquad \text{for every path } x_0, \dots, x_m \text{ with } m \le T_n \text{ that avoids } V \tag{2.2}$$

(Lemmas 10.4 and 12.3 and Proposition 12.2). Given (2.2), a chain with $R(v(X_0), X_k) \le 0$ for some $k \le T_n$ visits V at one of the times $0, \dots, k$. Since X_a has law π for every deterministic a, a union bound implies that

$$\mathbb{P}\big\{R(v(X_0), X_k) \le 0\big\} \le (k+1)\pi(V) \le 3(k+1)e^{-4L}$$

(see (12.9)). This is the only place where we use stationarity, and we use it only at deterministic times. For this reason the map v and the set V must be fixed before the chain is run, and they may not depend on its trajectory. In continuous time, the number of updates up to time e^L is a Poisson variable with mean ne^L, and by Markov's inequality it exceeds T_n with probability at most $1/n$. We conclude that

$$\pi\big\{x : d^J(x, e^L) \le \tau\big\} \le \frac{3(T_n+1)e^{-4L} + 1/n}{1/2 - \tau}$$

for every disorder in $\mathcal{G}_n$ (Proposition 12.5), and the right side tends to zero. When we integrate over the disorder, $\mathbb{P}(\mathcal{G}_n^c)$ is added once. The threshold τ enters only through the factor $1/(1/2 - \tau)$, which is why one coefficient serves every fixed $\tau < 1/2$ and both clocks.

2.2 Overlap constraints and reference configurations

We follow the architecture of the proof of [77, Theorem 1.1], which proves a barrier at the time $\exp(n^{1/10000})$ and the threshold $\tau = 1/4$.

Overlap constraints. For $(v, x, \widetilde{v}) \in \Sigma_n^3$, we write

$$r = \frac{R(v,x) + R(v,\widetilde{v})}{2}, \qquad \delta = \frac{R(v,x) - R(v,\widetilde{v})}{2}, \qquad q = R(x,\widetilde{v}),$$

so that the overlap matrix of $(v, x, \widetilde{v})$ is the matrix $D(r, q, \delta)$ of (4.16). We use two kinds of constraints. The constraints of the first kind act at fixed scales. For fixed cutoffs $0 < b_2 < b_1 < 1$ (Table 10.1), they state that if $R(x, \widetilde{v}) \ge q_*$ and $|R(v,x)|, |R(v,\widetilde{v})| \le b_1$,

then $|R(v,x) - R(v,\widetilde{v})| \le 2\eta_n$, and similarly with the cutoffs (b_1, b_2) in place of (q_*, b_1). The triples that violate them form the sets $\mathcal{B}_n(q_*, b_1)$ and $\mathcal{B}_n(b_1, b_2)$ of (10.18). In words, a large overlap between x and $\widetilde{v}$ forces their small overlaps with v to agree. The constraints of the second kind act at the scale ρ, near a level r in a compact interval $J_\beta \subset (0, q_\beta)$. With $\varepsilon = E\rho$ and $E = 1/100$, they state that no triple satisfies

$$r \in J_\beta, \qquad \varepsilon \le |\delta| \le 2\varepsilon, \qquad q \ge r + \rho.$$

Such triples are called narrow, and they form the set $\mathcal{N}_n$ (Section 10.2). Equivalently, if $R(x,\widetilde{v}) \ge r + \rho$, then $|R(v,x) - R(v,\widetilde{v})|$ does not lie in $[2\varepsilon, 4\varepsilon]$. We exclude a band, rather than a half-line, for the following reason. Along a path, every overlap changes by at most $2/n < 2\varepsilon$ in one step, and the difference $|R(v,x_k) - R(v,\widetilde{v})|$ cannot pass from below 2ε to above 4ε without lying in the band at some time. As long as the path keeps a large overlap with $\widetilde{v}$, a narrow constraint keeps its overlap with v within 2ε of $R(v,\widetilde{v})$ (Lemma 11.8).

Constrained pressures. In Sections 4 to 9, we write the three configurations as (v,x,y). For a feasible overlap matrix D, the probability that three independent samples from π have overlap matrix D is $Z_{3,n}(D)/Z_n^3$, where $Z_{3,n}(D)$ is the partition function restricted to the triples with overlap matrix D (Definition 4.16). There are at most $(n+1)^3$ feasible matrices, and it suffices to bound each numerator from above and the denominator from below. Upper bounds for the numerators come from Guerra's hierarchical interpolation [59] and its extension by Talagrand to replicas with prescribed overlaps [103, Section 3], [105, Propositions 3.4 and 4.4]. We use it in the form of [77, Proposition 2.3], stated as Proposition 4.19. For every admissible path (Q,m) of 3×3 covariance matrices, with positive semidefinite increments and nondecreasing masses, the comparison with zero terminal multiplier bounds the constrained pressure $p_{3,n}(D) = n^{-1}\mathbb{E}\log Z_{3,n}(D)$ by the trial value

$$p_{3,n}(D) \le \mathcal{T}_D(Q,m,0) = \Psi(Q,m,0) - \frac{\beta^2}{4}\int m\,\mathrm{d}\|Q\|_F^2 + \frac{\beta^2}{4}\|D - Q_N\|_F^2$$

(Definition 4.15 and Proposition 4.19). The last term is the cost of an endpoint Q_N that differs from D.

The baseline is the block hierarchy of [77, Section 4.2 and Section 7, Step 1], described in Lemma 4.20 and built from the Parisi distribution function $\alpha = \alpha_\beta$. Below the level r, the three fields share their increments, at mass $\alpha/3$. Between r and q, the fields of x and y share their increments, at mass $\alpha/2$, while the field of v evolves alone, and above q the three fields evolve independently. The endpoint of this path is $D(r,q,0)$, and its trial value is exactly $3P_\beta$. A perturbation of the baseline with endpoint $D(r,q,\delta)$ and a smaller trial value proves a pressure deficit. Whether such a perturbation exists is decided by the scalar Parisi problem. Let u be the spatial derivative of the solution of the Parisi

equation, let Y be the associated diffusion, and let $\Gamma(t) = \mathbb{E}u(t, Y_t)^2$. It was proved in [77, Proposition 3.2] that $\Gamma(q) = q$ at every point q of the support of the Parisi measure, and by [69, Theorem 1.1], the support is $[0, q_\beta]$. We conclude that $\Gamma(t) = t$ on $[0, q_\beta]$ (Corollary 4.8). For two configurations, we have the pair bound $p_{2,n}(D_2(q)) \le 2P_\beta - (\Gamma(q) - q)^2/2$ of [77, Lemma 4.3] (Proposition 4.23). This deficit vanishes on the support, and constraints at overlaps inside $(0, q_\beta)$ require the three-replica perturbations of [77, Sections 4 and 7] and of Section 7.

Reference configurations and crossings. In the dynamical argument, v and $\widetilde{v}$ are reference configurations and x is a configuration of the path. We take the references from lists of independent Gibbs samples, which are fixed before any path is considered (Section 10.3). Suppose that a path $x_0, \dots, x_m$ that avoids V satisfies $R(v(x_0), x_m) \le 0$. We proceed through levels $i = 0, 1, \dots, N_{\mathrm{lev}} - 1$, with thresholds $r_0 < r_1 < \cdots$ in an interval $[t_0, 2t_0] \subset J_\beta$ (Definition 11.2). At level 0, the reference is $v_0 = v(x_0)$, and k_0 is the first time with $R(v_0, x_k) \le r_0$. At level $j \ge 1$, a bound on the Gibbs mass of overlap neighborhoods supplies a new reference v_j with $R(v_j, x_{k_{j-1}}) \ge q_*$ (Lemma 10.1, which extends [77, Proposition 6.1]). We then search backward in time from k_{j-1} for the last time k_j with $R(v_j, x_k) \le r_j$. During the search, the constraints keep the overlaps of the path with the earlier references close to their values at k_{j-1}. The overlap with v_0 stays close to $r_0 < q_*$, while $R(v_0, x_0) \ge q_*$, so that the search finds the crossing before it reaches time 0 (Lemma 11.9). After the last level, at the single time $k_{N_{\mathrm{lev}}-1}$, the overlaps of the path with many references lie close to their thresholds. The thresholds are drawn at random, and a count over the time index shows that this is impossible when $m \le T_n$.

As in [77, Sections 8.3 and 8.4], the references are arranged in blocks of B levels. Within a block, a new reference has small overlaps with the earlier references of the block, and it deletes every retained reference with which its overlap exceeds b_1 in absolute value. Such conflicts occur only at the first level of a block, and they never reach the first $B - H \ge 2$ levels of the preceding block (Lemma 11.6).

2.3 The exponent

Three requirements fix the scales, which are listed in Table 2.2. The order in which the parameters are chosen is recorded in Tables 10.1 and 12.1.

The denominator. On the disorder event of Section 9, we have $F_n \ge n(P_\beta - c_0\rho^3)$, so that $Z_n^3 \ge e^{3n(P_\beta - c_0\rho^3)}$. To obtain this bound from Chebyshev's inequality with a fixed margin, we choose constants $a_0 < c_0$ such that $\Delta_{n,\beta} \le a_0\rho^3$ for all large n. Since $\mathbb{E}F_n = n(P_\beta - \Delta_{n,\beta})$, the distance from the mean of F_n to the threshold is then at least $n(c_0 - a_0)\rho^3$. The bounds on the numerators keep a deficit $\varkappa g\rho^3$, where $g\rho^3$ is a lower bound, uniform over the band, for the deficit per spin of the three-replica pressure near

Table 2.2. The scales of the proof of Theorem 1.1 and their roles. Here R, $E = 1/100$, S, and A_β are constants fixed before $n \to \infty$. The last five rows are the rates of decay of the error terms of Lemma 9.7, without constants and polynomial factors; see Tables 9.1 and 10.2.

Quantity	$\lambda = 7/30$	Role
ρ_n	$Rn^{-7/30}$	gap $q - r$ of the narrow constraints; length of a threshold interval
$\varepsilon_n = E\rho_n$	$ERn^{-7/30}$	overlap imbalance $\varepsilon_n \le \lvert\delta\rvert \le 2\varepsilon_n$; radius of the windows
η_n	$n^{-5/21}$	tolerance of the fixed-scale constraints
$s_n = S\rho_n$	$SRn^{-7/30}$	order of the fractional moment
L_n	$A_\beta n^{7/30}$	logarithmic time scale; logarithm of the list sizes
$n\rho_n^3$	$R^3 n^{3/10}$	Gibbs probability of narrow triples
$n\eta_n^2$	$n^{11/21}$	Gibbs probability of fixed-scale triples
$n\rho_n^4$	$R^4 n^{1/15}$	disorder failure of a local or separated-overlap bound
$n\eta_n^4$	$n^{1/21}$	disorder failure of a fixed-scale bound
$n^2\rho_n^6/n^{7/15}$	$R^6 n^{2/15}$	disorder failure of the denominator (polynomial)

a level r (Sections 2.4 and 2.5) and $\varkappa \in (0,1)$ is fixed. The ratio $Z_{3,n}(D)/Z_n^3$ is then at most $e^{-n(\varkappa g - 3c_0)\rho^3}$, which is small when $3c_0 < \varkappa g$. Constants $a_0 < c_0 < \varkappa g/3$ with these properties exist if and only if

$$\limsup_{n\to\infty} \frac{\Delta_{n,\beta}}{\rho^3} < \frac{\varkappa g}{3}$$

(see (S1)). With $\Delta_{n,\beta} \le C_b n^{-7/10}$, this holds for $\rho = Rn^{-7/30}$ if $R^3 > 3C_b/(\varkappa g)$ (see (10.15)). By Chebyshev's inequality and the bound $\operatorname{Var} F_n \le C_v n^{7/15}$ of Theorem 4.10, the lower bound on F_n fails with probability at most a constant times $n^{7/15}/(n^2\rho^6)$, which is $O(n^{-2/15})$. We note that Theorem 1.6 is not used in the proofs of Theorems 1.1 and 12.6.

The count. The thresholds lie in the intervals $I_i = [t_0 + 3i\rho, t_0 + (3i+1)\rho]$, of length ρ and spacing 3ρ. There are $N_{\mathrm{lev}} = MB$ levels, arranged in $M = \lfloor t_0/(3B\rho) \rfloor$ blocks (see (11.3)). In Section 2.6, we explain why a path of length m that avoids V and violates (2.2) produces, for every array of thresholds, a time $k \le m$ at which M overlaps lie in windows of radius $\Psi_0 = 2\varepsilon + 2\eta_n + 2/n$ around their thresholds (Corollary 11.11). For independent thresholds, uniform on their intervals, this event has probability at most

$$(m+1)\gamma_n^M, \qquad \gamma_n = \frac{2\Psi_0}{\rho} \longrightarrow 4E = \frac{1}{25} \tag{2.3}$$

(see (12.1) and (12.2)). If this quantity is less than one, such a path cannot exist. Note that M is of order $1/\rho$, so that we may take $\log m$ of order n^λ. This leads to the time horizon

e^L with the logarithmic time scale $L = An^\lambda$, for every $A < t_0 \log 25/(3BR)$ (see (12.4)). The count requires the ratio γ_n of the length of a window to the length ρ of a threshold interval to stay below one. With $\varepsilon = E\rho$, the overlap imbalance is of the largest permitted order, and this ratio stays below a fixed bound. The gap satisfies $q - r \geq \rho$, and the weakest local deficit occurs at the corner $q - r = \rho$, $|\delta| = E\rho$ of the band of Section 2.4.

In summary, the upper bound $n^{-3\lambda}$ on the mean bias allows $\rho \asymp n^{-\lambda}$, there are then of order n^λ levels, and the count allows $\log t$ of order n^λ: the exponent of the barrier is one third of the exponent of the upper bound on the mean bias. For the upper bound $n^{-7/10}$ of Theorem 4.11, this gives the exponent $7/30$ of Theorem 1.1.

The union bounds and the exponent $1/4$**.** The proof uses two kinds of union bound. In the disorder event, each of at most $(n+1)^3$ bounds on constrained partition functions near a level r fails with probability at most $e^{-cn\rho^4}$ (Section 2.5). The sufficient condition $n\rho^4/\log n \to \infty$ makes this polynomial union bound vanish (see (S3)). At a fixed disorder, we multiply a count e^{CL} of reference configurations and path positions by Gibbs probabilities bounded by $C'(n+1)^3 e^{-cn\rho^3}$, and it suffices that $cn\rho^3 - CL - 3\log(n+1) \to \infty$. Under the stronger condition $n\rho^3/L \to \infty$, we have this for every fixed C. For $\rho = Rn^{-\lambda}$ and $L = An^\lambda$, both sufficient conditions hold when $\lambda < 1/4$. At $\lambda = 1/4$, the quantity $n\rho^4 = R^4$ is bounded, and the polynomial union bound in the disorder event is ineffective, whether or not the conditional union bound can be absorbed by a choice of constants. The argument gives no exponent of $1/4$ or more, whatever the upper bound on the mean bias (Table 12.2). We note that a mean bias of order $n^{-3/4}$, as predicted by Franz, Parisi, and Ricci-Tersenghi [56], would correspond exactly to this threshold (Section 23).

2.4 The cubic deficit

We now describe the pressure deficit for the narrow constraints. We fix the compact intervals $J_\beta \subset \operatorname{int} I_\beta$ with $I_\beta \subset (0, q_\beta)$ of Section 10.2; the general statements of Sections 8 and 9 write them as $\mathcal{J} \subset \operatorname{int} I$. For a center $r \in J_\beta$, the gap $d = q - r$, and the overlap imbalance δ, we need a trial value for $D(r, q, \delta)$ below $3P_\beta$ by a fixed multiple of ρ^3, uniformly over $r \in J_\beta$ and over the band $\{\rho \leq d \leq d_0,\ E\rho \leq |\delta| \leq 2E\rho\}$. The local cutoff d_0 is chosen so small that every local window lies in I_β.

The obstacle. In [77, Section 7], the field of x receives δ times a Gaussian combination of the field increments in a window $[r-h, r+h]$, and the field of y receives $-\delta$ times the same combination. The resulting excess covariance in the antisymmetric direction $\mathbf{e}_x - \mathbf{e}_y$ contributes to the endpoint cost $(\beta^2/4)\|D - Q_N\|_F^2$, and the expansion of the trial value has the remainder $C(|\delta|^3 h^{-3/2} + \delta^4 h^{-2})$ (see (4.20)), which is small compared with the gain only if $|\delta|$ is much smaller than $h^{3/2}$, while h is smaller than $q - r$. In [77], $h = \rho^8$

and $|\delta|$ is of order ρ^{20}, and the resulting deficit is of order ρ^{43} [77, (7.32)]. When $|\delta|$ and $q - r$ are both of order ρ, however, the mismatch and the remainder are as large as the gain.

The observation. For the perturbation constructed below, at centers in the interior of the support, the terms of order at most two in the width parameter h of the window cancel exactly. The reason is the first stationarity identity $\beta^2 \mathbb{E}\chi(t, Y_t)^2 = 1$, where χ is the second spatial derivative of the Parisi solution. This identity is the derivative of $\Gamma(t) = t$, and it expresses the marginal stability of the Parisi solution [44, (22)–(23)]; its stochastic form is recorded in [69, Lemma 2.5]. The deficit is then of third order in h. To find it, we need a perturbed path whose endpoint is exactly $D(r, q, \delta)$, so that we incur no mismatch cost, and an expansion of its trial value to third order with a remainder of fourth order.

The perturbed path. We fix a ratio $\vartheta \in (0, 1/2]$; the dynamical argument uses $\vartheta = 1/2$. Given (r, d, δ), let $h = d/\vartheta$ be the width parameter of the window, $q_0 = r - h$ its left endpoint, and $k = \delta/h$ the relative imbalance (see (7.1)). The perturbed path has the five stages of Table 7.2. In the baseline, the first four stages are the following: the three fields share their increments on $[q_0, r]$, the fields of x and y share theirs on $[r, q]$ while v evolves alone, and all three are independent on $[q, q_0 + 2h]$. To perturb them, we add $\delta\,\Xi$ to the field of x and subtract it from the field of y, where

$$\Xi = \frac{1}{h}\{\omega_0(g_1 - g_2 - g_x - g_y) + \omega_1 g_v\}, \qquad \omega_0 = \frac{1}{2 - \vartheta}, \qquad \omega_1 = \frac{1 - \vartheta}{2 - \vartheta},$$

and g_1, g_2, g_v, g_x, g_y are the field increments of the window (see (7.4)). The weights are fixed, and the contrast extends past the time q at which x and y separate. By (7.5), the perturbation changes the covariance of the window in two ways: it shifts the overlaps of v with x and with y by $+\delta$ and $-\delta$, and it adds an excess covariance of order δ^2/h in the antisymmetric direction $\mathbf{e}_x - \mathbf{e}_y$. A fifth stage, the terminal correction, removes this excess exactly. It lasts a time proportional to hk^2, and it uses the positive semidefinite rate $\mathrm{Id} - (\mathbf{e}_x - \mathbf{e}_y)(\mathbf{e}_x - \mathbf{e}_y)^{\mathsf{T}}/4$ in place of Id. Every increment of the path is positive semidefinite, the masses are nondecreasing, and the endpoint is exactly $D(r, q, \delta)$ (Lemmas 7.1 and 7.2). From [77, Section 7] we keep the hierarchy in which the masses are divided by the sizes of the common blocks, the antisymmetric Gaussian contrast, and the source of the deficit in the identity $\beta^2 \mathbb{E}\chi(r, Y_r)^2 = 1$.

The reduction. We split the trial value $\mathcal{T}(q_0, h, k)$ into the initial segment $[0, q_0]$, on which the three fields share every increment, and the local window. Let $V(q_0, h, k, z)$ be the difference between the value of the local window on the diagonal $\{z_v = z_x = z_y = z\}$ and its unperturbed value $3\Phi(q_0, z)$ (see (7.11)), and let $\Delta\Pi$ be the change of the penalty $\int m\,\mathrm{d}\|Q\|_F^2$. The initial segment responds to the change V of its terminal function linearly,

up to an error $\|V\|_\infty^2/2$ that does not depend on the number of levels (Lemma 6.13 and Corollary 6.14). Since $\|V\|_\infty \le Ch^2k^2$, we have

$$\mathcal{T}(q_0,h,k) - 3P_\beta = \mathbb{E}V(q_0,h,k,Y_{q_0}) - \frac{\beta^2}{4}\Delta\Pi + O(h^4k^4)$$

(Corollary 7.5). The expectation is taken under the law of the scalar diffusion at the fixed time q_0, which does not depend on (h,k); this is why we expand at a fixed left endpoint. The local window is a composition of five short semilinear parabolic flows on $\mathbb{R}^3$, some with singular diffusion matrices. In Section 6 we bound their mixed derivatives in (h,k), uniformly in the field and including $h=0$ (Proposition 6.3). Using these bounds, we find that

$$V(q_0,h,k,z) = \frac{\beta^4\chi_0^2}{4}\Delta\Pi + \beta^6h^3\{\alpha(q_0)^2\chi_0^3\,C_\chi(k) + \alpha(q_0)\,w_0^2\,C_w(k)\} + O(h^4k^2) \qquad (2.4)$$

(Lemma 7.4 and Proposition 7.12). Here $\chi_0 = \chi(q_0,z)$ and $w_0 = w(q_0,z)$ are the second and third spatial derivatives of the Parisi solution, and C_χ and C_w are the trace polynomial and the Wick polynomial of (7.26) and (7.27). When the mass is frozen at $\alpha(q_0)$ on the window, the stages are Gaussian integration steps, and three formal steps with explicit coefficient matrices give the coefficients of order two and three (Lemmas 7.7 and 7.8 and Proposition 7.9). The variation of α on the window contributes, at order three, exactly the corresponding part of $\beta^4\chi_0^2\Delta\Pi/4$ (Lemma 7.11).

The cancellations. The penalty enters the trial value twice: directly, with the factor $-\beta^2/4$, and through V, with the factor $\beta^4\chi(q_0,Y_{q_0})^2/4$. After the expectation, the sum is $(\beta^2/4)(\beta^2\mathbb{E}\chi(q_0,Y_{q_0})^2-1)\Delta\Pi$, which vanishes by the first stationarity identity. Since $\Delta\Pi$ is of order h^2, this removes every term of order h^2, as well as the terms of order h^3 that involve $\alpha'(q_0)$. The second stationarity identity $\mathbb{E}w(t,Y_t)^2 = 2\alpha(t)\mathbb{E}\chi(t,Y_t)^3$ implies that the expectation of the braces in (2.4) equals $\alpha(q_0)^2\mathbb{E}\chi(q_0,Y_{q_0})^3\{C_\chi(k)+2C_w(k)\}$. We conclude that

$$\Big|\mathcal{T}(q_0,h,k) - 3P_\beta - \alpha(r)^2\beta^6\mathbb{E}\chi(r,Y_r)^3\,h^3P_\vartheta(k)\Big| \le Ch^4k^2, \qquad P_\vartheta = C_\chi + 2C_w \qquad (2.5)$$

(Proposition 7.13, with $q_0+h=r$). Both identities hold on $(0,q_\beta)$ (Lemma 5.3), and both were known; see [8, proof of Lemma 7.2] and [69, Lemma 2.5]. The use of the second identity in a covariance expansion is new. The polynomial P_ϑ is even, with coefficients $c_2(\vartheta) = 2\vartheta(\vartheta-1)/(\vartheta-2)^2 < 0$, $c_4(\vartheta) < 0$, and $c_6(\vartheta) > 0$ (see (7.33)). For $\vartheta = 1/2$,

$$P_{1/2}(k) = -\frac{2}{9}k^2 - \frac{62}{243}k^4 + \frac{6952}{19683}k^6$$

(see (7.34)), and Section 7.3 lists the traces from which this polynomial can be recomputed by hand.

The deficit on the band. In the variables d and δ, we have $-h^3 P_\vartheta(k) = \mathcal{G}_\vartheta(d, |\delta|)$, where

$$\mathcal{G}_\vartheta(d,x) = -\frac{c_2}{\vartheta}\,x^2 d - \vartheta c_4 \frac{x^4}{d} - \vartheta^3 c_6 \frac{x^6}{d^3}$$

(see (7.36)). The leading term is $2(1-\vartheta)(2-\vartheta)^{-2}\delta^2 d$, which is $4\delta^2 d/9$ for $\vartheta = 1/2$. The deficit is quadratic in the overlap imbalance and linear in the gap $q - r$; its minimum on the band is a positive multiple of ρ^3, and it can be larger away from the corner. The scalar factor $\alpha(r)^2\beta^6 \mathbb{E}\chi(r, Y_r)^3$ is bounded below on I_β by a constant $m_\beta > 0$, since $\alpha > 0$ on $(0, q_\beta]$ and $\mathbb{E}\chi(r, Y_r)^3 \geq \beta^{-3}$ (Corollary 5.5). For $E = 1/100$, the function $\mathcal{G}_\vartheta$ is increasing in d and in $|\delta|$ on the band, and it attains its minimum at the corner $(\rho, E\rho)$ (Lemma 7.14). Choosing d_0 so small that the remainder in (2.5) is absorbed, we obtain, for every fixed $g' < m_\beta E^2 G(\vartheta, E)$,

$$p_{3,n}\big(D(r,q,\delta)\big) \leq \mathcal{T}(q_0,h,k) \leq 3P_\beta - g'\rho^3 \tag{2.6}$$

for every feasible matrix in the band, all n, and all $0 < \rho \leq d_0$, uniformly over $r \in J_\beta$ (Proposition 7.15). Here $G(\vartheta, E)$ is the value of $E^{-2}\rho^{-3}\mathcal{G}_\vartheta$ at the corner, and $G(1/2, 1/100) > 1/3$. Since the second stationarity identity fails at $t = q_\beta$, where the Parisi measure has an atom (Remark 5.4), the centers lie in compact subintervals of $(0, q_\beta)$; no other condition on the Parisi measure near r is needed, in contrast with [77, Lemma 7.1].

2.5 From pressures to Gibbs probabilities

The obstacle. The bound (2.6) concerns the quenched pressure $n^{-1}\mathbb{E}\log Z_{3,n}(D)$. However, Gaussian concentration of this Lipschitz function of the disorder (Remark 4.22) turns a mean deficit of order ρ^3 only into the probability bound $e^{-cn\rho^6}$, and $n\rho^6$ tends to zero; for $\rho = Rn^{-7/30}$, it is $R^6 n^{-2/5}$.

Fractional moments. To overcome this obstacle, we bound positive fractional moments of order $s = S\rho$, for a small constant $S > 0$. Suppose that

$$\frac{1}{ns}\log \mathbb{E} Z_{3,n}(D)^s \leq 3P_\beta - g\rho^3. \tag{2.7}$$

Applying Markov's inequality to $Z_{3,n}(D)^s$, we find that

$$\mathbb{P}\big\{n^{-1}\log Z_{3,n}(D) > 3P_\beta - \varkappa g\rho^3\big\} \leq e^{-(1-\varkappa)gSn\rho^4},$$

and $n\rho^4$ grows like a power of n (Step 2 of the proof of Lemma 9.7). To prove (2.7), we adapt the matrix comparison of [77, Proposition 2.3] to moments of order s, as in the positive-moment interpolation of Aronow and Lopatto [8, Section 2.3]: the disorder is integrated at the outer mass s, instead of by an expectation. In this way we find

that the trial value of every admissible path whose masses are all at least s bounds $(ns)^{-1}\log \mathbb{E}Z_{\ell,n}(D)^s + \beta^2\ell^2 s/(4n)$ (Proposition 8.3). The interpolation is carried out for the completed Hamiltonian $H_n + G$ of Section 3, whose covariance is exactly $nR(x,y)^2/2$. The correction $\beta^2\ell^2 s/(4n)$ is quadratic in the number ℓ of configurations, because the same variable G enters every configuration (Lemma 3.1). Since this correction is nonnegative, it can be discarded.

The profile $\max\{\alpha, 3s\}$**.** The masses of the perturbed path on its initial segment are $\alpha(t)/3$, and $\alpha(t) \to 0$ as $t \to 0$, so that the requirement that all masses are at least s fails near the origin. To overcome this, we replace α by $\max\{\alpha, 3s\}$ on the initial segment. By the trial bound of Aronow and Lopatto [8, Lemma 7.2], the Parisi functional of this profile exceeds P_β by at most Cs^5 (Proposition 4.12 and Lemma 8.5), which is much smaller than ρ^3. The profile differs from α only on an interval $[0, Cs]$, far from the window, which starts at $q_0 \geq \min I_\beta > 0$. We show that its effect on the response of the initial segment to V is at most $C\|V'\|_\infty \|\max\{\alpha, 3s\} - \alpha\|_{L^1} \leq C\delta^2 s^2/h$ (Lemma 8.6). On the band this is at most a constant times $S^2\rho^3$, and we make it small by the choice of S (see (8.24)). Together with (2.6), this proves (2.7) for every fixed $g < m_\beta E^2 G(\vartheta, E)$, for all n and all small ρ, uniformly over the band (Lemma 8.7).

The other regimes. Every narrow triple falls into one of three regimes (Lemma 9.6). In the local regime $q - r \leq d_0$, we use (2.7). In the regime $q - r > d_0$ of separated overlaps, with $|q - \Gamma(q)|$ small, we use the quadratic deficit $c_{\mathrm{sep}}\delta^2$ of Lemma 9.5, which we derive from [77, Section 7, Steps 1–7] with fixed parameters; its source is the decay $\beta^2\mathbb{E}[\chi(q, Y_q)\chi(r, Y_r)] \leq 1 - c(q - r)$ of the mixed susceptibility [77, (7.31)]. Since δ^2 is of order ρ^2, Gaussian concentration suffices there. In the remaining pair regime, the two configurations with the larger overlap q have $|q - \Gamma(q)| > \nu$ for a fixed $\nu > 0$, so that the pair deficit $(\Gamma(q) - q)^2/2$ of [77, Lemma 4.3] (Lemma 9.1) applies. For the fixed-scale constraints, we use the three-replica deficit $c_{\mathrm{fix}}\delta^2$ of [77, (4.20)] (Lemma 9.2) at the tolerance $|\delta| \geq \eta_n$. In the separated-overlap, pair, and fixed-scale regimes, the mechanisms are those of [77]; what changes are the scales at which the pressure deficits are converted into probability bounds.

One disorder event. The disorder event $\mathcal{A}_n$ of Lemma 9.7 requires the lower bound on F_n of Section 2.3 and upper bounds on all constrained partition functions of the local, separated-overlap, fixed-scale, and pair families; these are deterministic families of at most $C(n+1)^3$ matrices. The event $\mathcal{A}_n$ fails with probability at most

$$C\frac{n^{7/15}}{n^2\rho^6} + C(n+1)^3\big(e^{-cn\rho^4} + e^{-cn\eta_n^4}\big) + C(n+1)e^{-cn}$$

(see (9.20)), which is $O(n^{-2/15})$ for $\lambda = 7/30$. For every disorder in $\mathcal{A}_n$, dividing the numerators by the denominator, we find that

$$\pi^{\otimes 3}(\mathcal{N}_n) \le C(n+1)^3 e^{-cn\rho^3}, \qquad \pi^{\otimes 3}(\mathcal{W}_n) \le C(n+1)^3 e^{-cn\eta_n^2},$$

where $\mathcal{W}_n$ is the set of fixed-scale triples (see (9.21)). We keep two kinds of error apart (Remark 9.9). Our estimate for the disorder failure probability $\mathbb{P}(\mathcal{A}_n^c)$ contains a polynomial term, and we use this estimate once, additively, at the end of the proof. The Gibbs probabilities at a fixed disorder in $\mathcal{A}_n$ decay at least as fast as $e^{-cn^{1-3\lambda}}$, and only these are multiplied by the numbers $e^{O(L)}$ of reference configurations and path positions. This separation replaces the conversion [77, Lemma 4.2] of pressure bounds into probability bounds, whose finite-size error $n^{-1/24}$ is far too large at the scale ρ^3.

2.6 The dynamical argument

Fixed families of references. The dynamical argument starts from two inputs at a fixed disorder. The first is the bound on the Gibbs mass of overlap neighborhoods: for every $S \subseteq \Sigma_n$, all $x \in S$ outside a set of Gibbs mass less than $e^{-D_1 L}$ have Gibbs mass at least $e^{-D_3 L}$ of configurations $z \in S$ with $|R(x,z)| \ge q_*$. We prove it by the argument of [77, Proposition 6.1], at every scale $L \to \infty$ with $L = o(n)$ (Lemma 10.1). The second input is the bound $\pi^{\otimes 3}(\mathcal{B}_n) \le q_n$ for the exceptional triples $\mathcal{B}_n = \mathcal{N}_n \cup \mathcal{B}_n(q_*, b_1) \cup \mathcal{B}_n(b_1, b_2)$ on the event $\mathcal{A}_n$, where $q_n = C(n+1)^3(e^{-cn\rho^3} + e^{-cn\eta_n^2} + e^{-cn})$ satisfies $e^{CL} q_n \to 0$ for every fixed C (Lemma 10.2).

In [77, Section 8.3], the properties of the lists of references are proved with probability tending to one jointly over the disorder, the lists, and the path, with bounds averaged over the disorder. At our scales, such an averaged bound contains the disorder failure probability, which does not remain small after multiplication by $e^{O(L)}$. For this reason, we fix the disorder first. At each disorder in $\mathcal{G}_n$, we sample the lists, with $\lceil e^{(c_s+1)L} \rceil$ independent Gibbs samples at the levels of position s in a block. The expected Gibbs mass of the configurations at which some cell contains no suitable entry is at most $2e^{-4L}$. The expected number of exceptional triples of list entries, and the expected Gibbs mass of the configurations that form one with two entries, are at most $e^{O(L)} q_n$. We fix a realization at which the sum of such quantities is at most its mean (Lemma 10.4). Outside a set V with $\pi(V) \le 3e^{-4L}$, it has properties (P1)–(P4) of Section 11.1: every cell supplies a new reference for each configuration in it, the fixed-scale constraints hold across q_* and across b_1, and no configuration outside V forms a narrow triple with two signed entries. The family and V depend only on J (Remark 10.5).

The accumulation of errors. In [77, (8.19)–(8.20)], the overlaps of every retained reference with the current configuration and with the later references are controlled within $6j\varepsilon$ after

j levels. These errors accumulate, and the final count of [77, (8.29)] requires $N_{\mathrm{lev}}\varepsilon$ to be much smaller than ρ. Since N_{lev} is of order $1/\rho$, this forces ε to be much smaller than ρ^2. Our choice $\varepsilon = E\rho$, at which the narrow constraints have a deficit of order ρ^3, is far outside this range.

Next retained references. To overcome this, we compare each retained reference v_i with the next retained reference $v_{a_j(i)}$, where $a_j(i)$ is the smallest retained level larger than i after level j (Section 11.2). We call the overlap $c_j(i) = R(v_i, v_{a_j(i)})$ the center of v_i. When v_{i+1} is selected at the time k_i, we have $R(v_{i+1}, x_{k_i}) \geq q_*$ and $R(v_i, x_{k_i}) \in [r_i - 2/n, r_i]$, and the fixed-scale constraint across q_* implies that $|c_j(i) - r_i| \leq 2\eta_n + 2/n$. The center changes only when a deletion removes the next retained reference of i. The next retained reference is then the first reference of the following block, which is never deleted, and the center moves by at most $2\varepsilon + 2\eta_n$. We show by induction that, for every retained $i < j$,

$$|c_j(i) - r_i| \leq 2\varepsilon + 4\eta_n + \frac{2}{n}, \qquad |R(v_i, x_{k_j}) - c_j(i)| < 2\varepsilon$$

(Definition 11.4 and Proposition 11.10). The errors are of order ε at every level, and they do not grow with the number of levels.

The narrow constraints preserve these windows during the backward search (Lemma 11.8). Suppose that some window fails at a time between k_j and k_{j-1}, and let k be the largest such time and v_i a reference whose window fails at k. All windows hold at time $k+1$, and since one step changes overlaps by at most $2/n$, the overlap $R(v_i, x_k)$ differs from $c_j(i)$ by an amount in $[2\varepsilon, 2\varepsilon + 2/n)$. Let $a = a_j(i)$. Then $R(v_a, x_k)$ is close to r_a or larger: if $a = j$, because the backward search has not yet reached the crossing of r_j, and otherwise because the window of v_a holds at time $k+1$. Since $r_a - r_i \geq 2\rho$ and all errors are of order ε, we find that $R(v_a, x_k) \geq r + \rho$, where r is the mean of $R(v_i, x_k)$ and $R(v_i, v_a) = c_j(i)$. The triple (v_i, x_k, v_a) is then narrow, which (P4) excludes because $x_k \notin V$. In other words, the next retained reference supplies the large overlap that the narrow constraint needs, and the spacing 3ρ of the threshold intervals makes this overlap large enough.

The count. At the time $k_* = k_{N_{\mathrm{lev}}-1}$, the first reference $v_{\ell B}$ of each block ℓ still has its original next retained reference $v_{\ell B+1}$, because the first two levels of a block are never deleted. By its window, we have $|R(v_{\ell B}, x_{k_*}) - r_{\ell B}| < \Psi_0$ for $\ell = 0, \dots, M-1$ (Corollary 11.11). The reference of each level depends only on the thresholds of the earlier levels (Lemma 11.3); this is why a selected reference whose crossing fails is kept. Removing the windows one at a time, from the largest level down, we find that for each fixed time k the M windows hold with probability at most γ_n^M over independent uniform thresholds (Lemma 12.1). Summing over $k \leq m$, we have the bound (2.3). If $(m+1)\gamma_n^M < 1$, some threshold array admits no such time, which contradicts the existence of a path that violates (2.2) (Proposition 12.2). In [77, (8.29)], the count is a

union over subsets of retained levels; here the windows sit at the deterministic first levels of the blocks, and we count only M of them. Further, since the static estimates are uniform over the interval J_β, we need no random event that restricts the thresholds, such as [77, (8.12)].

Quantitative and almost-sure forms. With the lower bound on F_n from the small-deviation inequality of Paouris and Valettas [90, Theorem 2.1 and Remark 2.4(3)] instead of Chebyshev's inequality, with $\eta_n = \rho_n/1000$, and with $A < t_0 \log(250/11)/(3BR)$, the disorder event fails with probability at most $Ce^{-cn^{1/15}}$, and on it the initial configurations from which the distance stays at least $1/2 - Ce^{-cn^{7/30}}$ up to the time horizon $e^{An^{7/30}}$ have Gibbs mass at least $1 - Ce^{-cn^{7/30}}$, for all $n \geq n_1(C_{\mathrm{rate}})$ and every single-site generator with rates at most a fixed C_{rate} that leaves π stationary (Theorem 12.6). By the same theorem, we have the relaxation time bound $e^{An^{7/30}}/(Cn^{7/30})$ for the reversible ones and, by the Borel–Cantelli lemma, the almost sure form for every coupling of the disorders of different sizes.

2.7 The upper bound

We follow the architecture of the proof of [85, Theorem 1], which proves Theorem 1.3 for the saving $\omega = 1/40000000$. The proof has four steps. A finite-chain argument reduces a bound for typical initial configurations to a lower bound on the stationary flow across every cut whose two sides both have appreciable Gibbs mass. A random disorder and a random configuration are then constructed jointly, by Gaussian queries of the disorder matrix that follow the scalar Parisi diffusion. The law of this pair is compared with the Gibbs joint law in two ways: through its relative entropy, and through an upper bound on its likelihood ratio outside a rare event. A Gaussian smoothing of the disorder reverses the entropy comparison, so that the constructed configuration reaches every set of appreciable Gibbs mass. Finally, a sequence of such configurations, built from rotated Gaussian inputs, connects the two sides of a cut of small flow while avoiding its boundary, which is impossible. In [85], the parameters are fixed numbers, and the exponent is not optimized.

To reach every $\omega < 1/4$, two estimates must be close to sharp: the relative entropy of the proposal at the longer horizon, and the likelihood bound at the shorter horizon, together with the probability of its exceptional event. The first is described in the paragraph on the forward entropy below, and the second in the paragraph on coarse domination. The other steps follow [85], with the changes of parameters that these two estimates allow.

Notation. Let $g \in \mathbb{R}^{n\times n}$ have independent standard Gaussian entries, with law γ_n, and let $W(g) = (g + g^{\mathsf{T}})/\sqrt{2n}$. With $J_{ij} = (g_{ij} + g_{ji})/\sqrt{2}$ and $J'_{ii} = g_{ii}$, the matrix $W(g)$ is

Table 2.3. The exponents of the proof of Theorem 1.3, as functions of $\varepsilon = \varepsilon_\omega = (1/4 - \omega)/20$, their values for $\omega = 1/5$, and their roles. They are fixed after β and ω, and before $n \to \infty$; see Sections 13.3 and 19.5.

Exponent	Value	$\omega = 1/5$	Role
a	$1/2$	$1/2$	mesh $h = n^{-\mathsf{a}}$
f	$1/2$	$1/2$	fine horizon $V_{\mathsf{f}} \approx \mathsf{f} \log n$
b	$1/2 - 14\varepsilon$	$93/200$	coarse horizon $V_{\mathsf{b}} \approx \mathsf{b} \log n$
$\mathsf{k_f}$	$1/2 - 2\varepsilon$	$99/200$	forward entropy $n^{1-\mathsf{k_f}}$
D	$1/4 - \varepsilon$	$99/400$	smoothing $\delta = n^{-\mathsf{D}}$
$\mathsf{u_E}$	$1/4 - 2\varepsilon$	$49/200$	reverse entropy $n^{1-\mathsf{u_E}}$
d	$\varepsilon/2$	$1/800$	masses $n^{-\mathsf{d}}$ of the cuts
e	$1/4 - 3\varepsilon$	$97/400$	crossing probability $e^{-2n^{1-\mathsf{e}}}$
w	$1/8 - 3\varepsilon$	$47/400$	weighted errors $n^{-\mathsf{w}}$
z	$1/4 - 8\varepsilon$	$23/100$	accuracy $n^{-\mathsf{z}}$ of the row averages
$\mathsf{k_b}$	$1/4 - 9\varepsilon$	$91/400$	likelihood bound $e^{n^{1-\mathsf{k_b}}}$
$\mathsf{s_b}$	$1/4 - 5\varepsilon$	$19/80$	failure $e^{-n^{1-\mathsf{s_b}}}$ of the coarse proposal
$\mathsf{s_{seq}}$	$1/4 - 4\varepsilon$	$6/25$	failure $e^{-n^{1-\mathsf{s_{seq}}}}$ of the sequence
r	$1/4 - 15\varepsilon/2$	$37/160$	Hamming radius $R_n = \lceil n^{1-\mathsf{r}} \rceil$
$\mathsf{u_F}$	$1/4 - 39\varepsilon/4$	$361/1600$	threshold $n^{1-\mathsf{u_F}}$ for sums of local fields
q	$1/4 - 10\varepsilon$	$9/40$	flow threshold $\phi_n = e^{-n^{1-\mathsf{q}}}$
$\mathsf{z_P}$	$3/10$	$3/10$	pressure event $F^c \le nP_\beta + n^{1-\mathsf{z_P}}$
$\mathsf{g_{rot}}$	$3/5$	$3/5$	grid of $\ell_n + 1 \approx e^{n^{\mathsf{g_{rot}}}}$ angles

the matrix W of Section 3.2, and $H_g^c(x) = x^{\mathsf{T}} W(g) x/2$ is the completed Hamiltonian of the disorder $J(g)$. We write $F^c(g)$ for the completed log partition function, $\pi_g = \pi^{J(g)}$, $P_t^g = P_t^{J(g)}$, and

$$\Pi_n(\,\mathrm{d}g, x) = \gamma_n(\,\mathrm{d}g)\, \pi_g(x)$$

for the Gibbs joint law. Its $(J(g), x)$ marginal is the realized-start law $\mathbb{P}_{J, X_0 \sim \pi^J}$ of Definition 3.3. For $S \subseteq \Sigma_n$, the stationary flow across S is

$$\mathcal{Q}_g(S, S^c) = \sum_{x \in S,\, y \notin S} \pi_g(x)\, \mathcal{L}^{J(g)}(x, y),$$

where $\mathcal{L}^{J(g)}(x, x^{i,-x_i}) = p_i^{J(g)}(-x_i \mid x)$ is the rate of a flip at site i (Section 3.3). The exponents of the proof are fixed numbers, written in sans-serif ($\mathsf{a}, \mathsf{q}, \mathsf{k_f}, \dots$); they are chosen after β and ω, and they are listed in Table 2.3. Constants $C_\beta, c_\beta > 0$ depend only on β and on the exponents.

From balanced-cut flow to mixing. By Lemma 3.4, Theorem 1.3 states that the Gibbs mass of the configurations x with $d^{J(g)}(x, e^{n^{1-\omega}}) > 1/4$ tends to zero in probability over

g. The reduction to flows is the finite-chain lemma [85, Lemma 8], which we prove as Lemma 19.1. Let $0 < \epsilon < 1/3$ and $\phi > 0$, and suppose that $\mathcal{Q}_g(S, S^c) > \phi$ for every set S with $\epsilon \le \pi_g(S) \le 1 - \epsilon$. Then, for every $t > 0$,

$$\sum_{x \in \Sigma_n} \pi_g(x)\, d^{J(g)}(x,t) \le \frac{n\sqrt{2\log 2}}{\phi\sqrt{t}} + 2\epsilon. \tag{2.8}$$

Here n bounds the total jump rate out of a configuration, and $n \log 2$ bounds the entropy of π_g. The proof removes, one at a time, sets of at most half of the remaining mass whose flow to the rest is less than ϕ times their mass. The hypothesis prevents the removed mass from reaching ϵ. On the remaining set, the flow bound gives a Poincaré inequality, by the argument of Lawler and Sokal [67, Theorem 3.5], and the entropy dissipation of the full chain [51, Lemmas 2.5 and 2.7] bounds the average of the distances from fixed initial configurations. The trajectory need not stay in the remaining set. Initial configurations from which the chain is slow are allowed, and a set of them costs at most its Gibbs mass. For this reason only cuts of mass at least ϵ enter, and the conclusion concerns typical initial configurations; at large β, the worst-case mixing time is exponential [99, Theorem 1.1].

We apply (2.8) with $\epsilon = n^{-\mathsf{d}}$, $\phi = \phi_n = e^{-n^{1-\mathsf{q}}}$, and $t = e^{n^{1-\omega}}$. By Markov's inequality in the Gibbs draw, we find that the joint probability that $d^{J(g)}(x,t) > 1/4$ is at most

$$\gamma_n(\mathcal{W}_n) + 2e^{-n} + 4n\sqrt{2\log 2}\, e^{n^{1-\mathsf{q}} - n^{1-\omega}/2} + 8n^{-\mathsf{d}}$$

(Section 19). Since $\omega < \mathsf{q}$, the third term tends to zero. The term $2e^{-n}$ bounds the probability that $\|W(g)\|_{\mathrm{op}} > 12$ (Lemma 13.2), and

$$\begin{aligned}\mathcal{W}_n = \big\{ g : \|W(g)\|_{\mathrm{op}} \le 12, \text{ and there exists } S \subseteq \Sigma_n \\ \text{with } n^{-\mathsf{d}} \le \pi_g(S) \le 1 - n^{-\mathsf{d}} \text{ and } \mathcal{Q}_g(S, S^c) \le \phi_n \big\}\end{aligned}$$

is the event that a cut of small flow exists. The rest of the proof shows that $\gamma_n(\mathcal{W}_n) \to 0$ (Proposition 18.9).

The proposal. The flow bound is proved with a random pair (g, x) whose law is close to Π_n. Its construction is that of [85, Section 4], which is related to the iterative construction of solutions of the TAP equations by Bolthausen [25] and to the incremental approximate message passing algorithm of Montanari [75]. Let α_β be the Parisi distribution function; the Parisi measure has support $[0, q_\beta]$ (Theorem 4.7). In the variance time $v = \beta^2 t$, the scalar diffusion of Section 4 solves

$$\mathrm{d}Y_v = \mathrm{d}M_v + \bar{\alpha}(v)\, m(v, Y_v)\, \mathrm{d}v, \qquad Y_0 = 0,$$

where M is a standard Brownian motion, $\bar{\alpha}(v) = \alpha_\beta(v/\beta^2)$, and $m(v,y) = u_{\alpha_\beta}(v/\beta^2, y)$ is the spatial derivative of the Parisi solution. The magnetization $m_v = m(v, Y_v)$ is a

martingale, with $\mathrm{d}m_v = \chi_v \,\mathrm{d}M_v$, where $\chi_v = \chi(v, Y_v)$ is the second spatial derivative. Beyond $v_* = \beta^2 q_\beta$, we continue the process with $\bar{\alpha} = 1$, $m(v, y) = \tanh y$, and $\chi(v, y) = \operatorname{sech}^2 y$; the continued potential solves the same equation (Section 13.4). We define

$$p(v) = \mathbb{E}\chi_v, \qquad r(v) = \mathbb{E}\chi_v^2, \qquad \rho(v) = \beta\sqrt{r(v)}, \qquad \sigma(v) = \sqrt{1 - \rho(v)^2}.$$

By Corollary 4.8, we have $r = \beta^{-2}$ on $[0, v_*]$, so that $\rho = 1$ and $\sigma = 0$ there. After v_*, the functions p and r decay at the rate $e^{-v/2}$ up to a factor $\sqrt{1+v}$ (Proposition 13.4 and Lemma 13.6), and at a horizon V of order $\log n$ the magnetization is close to ± 1.

In the proposal, we run one copy of the continued scalar process for each coordinate, driven by noise extracted from the matrix $W = W(g)$. We cut the time axis into cells $[t_j, t_{j+1}]$ with

$$t_0 = 0, \qquad h_j = h e^{t_j/4}, \qquad t_{j+1} = t_j + h_j, \qquad h = n^{-\mathsf{a}}.$$

At the query $j \geq 1$, we take the increment of the vector of magnetizations over the previous cell, normalized as $\mathcal{T}_j = (m(t_j) - m(t_{j-1}))/s_j$ with $s_j^2 = \int_{t_{j-1}}^{t_j} r$, and we orthogonalize it against the earlier query directions in the inner product $n^{-1}\langle u, v\rangle$. This gives a direction u_j, and we read the column Wu_j. Its component orthogonal to $u_1, \dots, u_j$, together with an independent Gaussian completion inside their span, is a vector e_j that is standard Gaussian given the past. The vector e_j becomes the standardized increment of the query paths B_i over the next cell $[t_j, t_{j+1}]$, which is filled with independent Brownian bridges. Each coordinate is driven by $M_i = \int \rho \,\mathrm{d}B_i + \int \sigma \,\mathrm{d}D_i$, where the D_i are independent Brownian motions. Up to v_*, the matrix drives the process alone; after v_*, the query noise is mixed with independent noise. By Gaussian orthogonal invariance, the unread part of W is a compressed Gaussian orthogonal ensemble, independent of the past. Then the matrix assembled from the answers and the remaining Gaussian variables has the law of $W(g)$ under γ_n, independently of the initial data, and the pairs (B_i, D_i) are independent pairs of Brownian motions (Lemma 14.1). At the horizon V, the spin is linked to the terminal magnetizations by the density $\prod_i (1 + x_i m_i(V))$ relative to an independent uniform spin (Definition 14.2). The output Q_V is the law of (g, x), where g is obtained from W by adding an independent antisymmetric Gaussian matrix, so that $W(g) = W$. Its marginals are γ_n and the uniform law, but x is correlated with g.

The construction has two equivalent forms: in one, the matrix is assembled from the answers to the queries, and in the other, the same recursion reads a given matrix and given completions (Definition 14.2 and Lemma 14.1). The main structural fact is that, under the linked law, the rows (x_i, Y_i, B_i, D_i), $i = 1, \dots, n$, are independent, with an explicit one-row law ν_V. Apart from the Gaussian variables that the recursion never reads, the statistics of the proposal used below are averages of n independent rows, or explicit functions of such averages.

We use two horizons on the same mesh: the fine horizon V_{f} and the coarse horizon V_{b}, the first times t_j beyond $\mathsf{f}\log n$ and $\mathsf{b}\log n$, with $\mathsf{b} < \mathsf{f}$ (Lemma 14.3). The two runs read the same inputs, so that the coarse run is an initial segment of the fine run. The growing cells are new compared with [85], where the cells have equal length $n^{-1/16}$. Since $r(t_j)$ decays like $e^{-t_j/2}$, the lengths h_j grow as the profile decays, and $h_j^2 r(t_j)$ stays of order h^2 at most. The number of queries is at most $8n^{\mathsf{a}}$ for every horizon (Lemma 14.3), and the errors of the mesh add up to $O(h)$, uniformly in the horizon. Unequal cells require some care in the estimates: $\mathcal{T}_j$ is an increment over the previous cell, e_j belongs to the next cell, and the ratio $h_j/h_{j-1} = e^{h_{j-1}/4}$ enters the bounds.

The forward entropy. The first comparison is

$$\mathrm{KL}(Q_V \| \Pi_n) \le n j(V) - n\log 2 + \mathbb{E}F^c(g) - \beta\,\mathbb{E}H_g^c(x), \qquad j(V) = \frac{1}{2}\int_0^V \rho(v)^2 \varkappa(v)\,\mathrm{d}v, \tag{2.9}$$

where $\varkappa(v) = \mathbb{E}[\chi_v^2/(1-m_v^2)]$ and the last expectation is under the proposal (Lemma 15.2). The quantity $j(V)$ is the information per spin that the proposal acquires through the matrix. To see this, we move the part of the linking density that comes from the independent noise D into a reference law, as in [85, Section 4], whose (g,x) marginal is γ_n times the uniform law. The logarithm of the remaining density is a sum of n independent rows, each with mean $j(V)$ (Lemma 15.1). The factor ρ^2 is the share of the matrix in the driving noise. Changing the reference to Π_n, whose density is $2^n e^{\beta H_g^c(x) - F^c(g)}$, gives (2.9).

The terms of (2.9) are related by the scalar identity [85, Lemma 4]: for every $V \ge v_*$,

$$j(V) - \log 2 + P_\beta - \beta e(V) = \frac{\beta^2}{4} p(V)^2, \qquad e(V) = \beta\int_0^V r(v)p(v)\,\mathrm{d}v \tag{2.10}$$

(Proposition 13.4). Here $e(V)$ is the scalar reference for the energy per spin of the proposal, and the identity balances the information paid through the matrix against the energy gained. By Guerra's bound and Lemma 3.1, we have $\mathbb{E}F^c(g) \le nP_\beta$ (Theorem 4.4). The expected energy satisfies

$$\frac{1}{n}\mathbb{E}H_g^c(x) \ge e(V) - C_\beta\Big[h + \frac{(N+1)(1+V)+\sqrt{N}}{n} + \frac{(N+1)^2 V R_V}{n^2}\Big],$$

where N is the number of queries and $R_V = (\min_{[0,V]} r)^{-1}$ (Lemma 15.4). Inserting these bounds into (2.9), and using $p(V)^2 \le C_\beta e^{-V}$, we find that

$$\mathrm{KL}(Q_V \| \Pi_n) \le C_\beta\Big[nh + ne^{-V} + (N+1)(1+V) + \sqrt{N} + \frac{(N+1)^2 V R_V}{n}\Big].$$

At the fine horizon, the five terms have the powers $1-\mathsf{a}$, $1-\mathsf{f}$, a, $\mathsf{a}/2$, and $2\mathsf{a}+\mathsf{f}/2-1$ of n, up to logarithms, and we obtain $\mathrm{KL}(Q_F\|\Pi_n) \le C_\beta n^{1-\mathsf{k_f}}$ for the fine output Q_F whenever $\mathsf{a} > \mathsf{f}/4$, $\mathsf{a}+\mathsf{f}/2 < 1$, and $\mathsf{k_f} < \min(\mathsf{a}, \mathsf{f}, 1-\mathsf{a})$ (Proposition 15.5).

This is the first place where a sharp estimate is needed. In [85, Proposition 5], the relative entropy is at most $2n^{1-l/2}$ for $0 < l \le 1/100$; the energy is computed on the event where the empirical Gram matrix of the queries is close to the identity, and the information is concentrated separately. Here the expected energy is computed directly, through an exact conditional identity for the third moments of a single query (Lemma 15.3), and no Gram event enters. The saving $\mathsf{k_f}$ can then be taken arbitrarily close to $1/2$, and $1/2$ is the limit of the estimate: the error nh of the mesh and the number n^{a} of queries balance at $\mathsf{a} = 1/2$. For $\mathsf{a} = \mathsf{f} = 1/2$, we have $\mathrm{KL}(Q_F \| \Pi_n) \le C_\beta n^{1/2+2\varepsilon_\omega}$, where $\varepsilon_\omega = (1/4 - \omega)/20$.

Coarse domination. The second comparison is a bound on the likelihood of the coarse proposal. There exists an event E_B of the coarse proposal, with $\mathbb{P}(E_B^c) \le e^{-n^{1-\mathsf{s_b}}}$, such that the output restricted to E_B has a kernel $q_B^{\mathrm{good}}(g,x)$ with respect to γ_n that satisfies

$$q_B^{\mathrm{good}}(g,x) \le e^{-nP_\beta + n^{1-\mathsf{k_b}}} e^{\beta H_g^c(x)} = e^{n^{1-\mathsf{k_b}} + F^c(g) - nP_\beta}\, \pi_g(x) \tag{2.11}$$

for all x and almost every g (Proposition 17.8). When $F^c(g) \le nP_\beta + n^{1-\mathsf{z_P}}$, this is a bound by $e^{o(n^{1-\mathsf{q}})}$ times the Gibbs measure once $\mathsf{q} < \mathsf{k_b}$ and $\mathsf{q} < \mathsf{z_P}$. Since the log density of Π_n relative to γ_n times the uniform law is $n \log 2 + \beta H_g^c(x) - F^c(g)$, it suffices to show that, on E_B,

$$\frac{1}{n} I^{(n)} - \log 2 - \frac{\beta}{n} H_g^c(x) \le -P_\beta + n^{-\mathsf{k_b}}, \tag{2.12}$$

where $I^{(n)} = \sum_i I_{V,i}$ is the logarithm of the density of the proposal relative to the reference law of the paragraph on the forward entropy, at $V = V_{\mathsf{b}}$. Integrating (2.12) against the reference law, we obtain (2.11); the factor 2^n cancels the uniform spin weight 2^{-n}.

The obstacle is the accuracy. By (2.10), the left side of (2.12) has mean close to $-P_\beta$, and we must show that it deviates by at most $n^{-\mathsf{z}}$, for an exponent $\mathsf{z} > \mathsf{k_b} > \mathsf{q} > \omega$, outside an event of probability $e^{-n^{1-\mathsf{s_b}}}$, with $\mathsf{s_b} < 1/4$. If the information and the energy are concentrated separately, as in [85, Proposition 5], each row has variance of order one, and a deviation $n^{-\mathsf{z}}$ has probability of order $e^{-n^{1-2\mathsf{z}}}$. In the parameter system below, the resulting requirement $2\mathsf{z} < \mathsf{s_b}$, together with $\omega < \mathsf{q} < \mathsf{k_b} < \mathsf{z}$ and $\mathsf{s_b} < 1/4$ would confine ω below $1/8$. We avoid this loss with three observations.

The first is algebraic. We write $H_g^c(x)/n = E_{\mathrm{qry}} + Z$, where E_{qry} is the energy carried by the queried columns and Z collects the diagonal entries $\langle u_j, W u_j \rangle / n$ and the unread block. The queried energy is an explicit function of the empirical Gram and cross matrices of the vectors $\mathcal{T}_j$, e_j, and x, whose population values are known. Its linearization in the errors of these matrices is itself an average of independent rows: it equals $n^{-1} \sum_i \zeta_{h,i}$, where

$$\zeta_h = \sum_j w_j (A_{t_j} - A_{t_{j-1}}) + \sum_j s_j (x - m_{t_j}) e_j - \frac{e_h}{2}, \qquad A_v = x m_v - \frac{m_v^2}{2},$$

for a row with spin x, magnetization m, and query endpoints e_j. The deterministic weights are $b_j = h_j^{-1/2} \int_{t_j}^{t_{j+1}} \rho(v)p(v)\,\mathrm{d}v$ and $w_j = b_j/s_j$, and $e_h = \sum_j s_j b_j$ is the discrete scalar energy. With these normalizations, $\mathbb{E}\zeta_h = e_h$ (Lemma 17.3). The remainder of the linearization is quadratic in a weighted measure W_{wt} of the errors (Lemma 17.2). This gives

$$\frac{1}{n} I^{(n)} - \frac{\beta}{n} H_g^c(x) = \frac{1}{n} \sum_{i=1}^{n} \bigl(I_{V,i} - \beta\zeta_{h,i}\bigr) + O\bigl(D_* W_{\mathrm{wt}}^2\bigr) - \beta Z,$$

where $D_* = 1 + \lceil \log_2 N \rceil$.

The second observation is a pathwise cancellation between information and energy (Lemma 17.5). On a mesh of vanishing size, ζ_h becomes $\zeta_\infty = Z_V - e(V)/2$, where $Z_V = \beta \int_0^V p\,\mathrm{d}A + \int_0^V \sqrt{r}\,(x-m)\,\mathrm{d}B$. For every $V \geq v_*$, we have exactly

$$I_V - \beta\zeta_\infty = \log 2 - P_\beta - \frac{\beta^2}{4} p(V)^2 + \frac{\beta^2 p(V)}{2} (x - m_V)^2. \tag{2.13}$$

In other words, the stochastic integrals of the information and of the energy cancel along every path. Note that the expectation of (2.13) is (2.10). To prove (2.13), we use Itô's formula for the Parisi solution along the scalar process (Lemma 17.4). Before v_*, it shows that $\beta^2 pA + xY - U(v,Y)$ has differential $\beta\,\mathrm{d}Z_v$, where U is the potential, and that the driver B equals M; at v_*, the identity $\log(1 + x\tanh y) = xy - \log\cosh y$ matches the two sides; after v_*, the difference $I_v - \beta Z_v$ moves only through $\beta^2 p(v)(x-m_v)^2/2$. The last term of (2.13) does not have variance of order h: its centered version T_V has variance $\beta^4 p(V)^2\{4p(V) - 3r(V) - p(V)^2\}/4 \leq \beta^4 p(V)^3$, which is positive at every fixed V. Its amplitude is at most $2\beta^2 p(V) = O_\beta(n^{-\mathsf{b}/2})$, however, and it is absorbed deterministically when $\mathsf{z} < \mathsf{b}/2$.

The third observation concerns the difference between the mesh and its limit. The residual $R_h = Z_V - \zeta_h - e_h/2$ has mean $3(e(V) - e_h)/2 = O_\beta(h)$, and its centered moment generating function is at most $e^{C_\beta \theta^2 h}$ for $|\theta| \leq c_\beta/h$, uniformly in the horizon and in the number of cells (Lemma 17.6). On the growing mesh, the lag of the magnetization inside a cell of length h_j contributes $h_j^2 r(t_j)$, which is at most of order h^2, and there are at most $8/h$ cells. Since the rows are independent, we conclude that the average of $I_{V,i} - \beta\zeta_{h,i}$ exceeds $\log 2 - P_\beta + C_\beta n^{-\mathsf{z}}$ with probability at most $2e^{-c_\beta n^{1+\mathsf{a}-2\mathsf{z}}}$ (Lemma 17.7). The loss is $2\mathsf{z} - \mathsf{a} < 0$, in place of $2\mathsf{z}$.

It remains to make W_{wt}^2 smaller than $n^{-\mathsf{z}}$ and to control Z. We have $W_{\mathrm{wt}} \leq C_\beta n^{-\mathsf{w}} \log^{\mathsf{C}} n$ except with probability

$$2\exp\bigl\{C_\beta n^{\mathsf{a}} - c_\beta n^{1-\max\{2\mathsf{w},\mathsf{w}+\mathsf{b}/4\}} / \log^{\mathsf{C}} n\bigr\}$$

(Lemma 16.6). The Gram–Schmidt step is a Cholesky factorization, and its errors involve the triangular truncation $\mathrm{Up}_{1/2}$, which keeps the strict upper triangle and half of the

diagonal. As in [85, Section 4], a dyadic decomposition of the triangle costs only powers of D_*. The quantity W_{wt} measures the errors through their projections onto dyadic intervals I of query indices, weighted by the population vectors. These projections are built from martingale increments over the previous cells, and their variances are bounded by a polylogarithmic multiple of the total length $d_I = \sum_{j \in I} h_{j-1}$ of these cells. We divide each projection by $\sqrt{d_I}$ before applying a variance-sensitive Bernstein inequality. Then the variance parameter is polylogarithmic, and the inverse profile $R_V \le C_\beta n^{\mathsf{b}/2} \sqrt{\log n}$ enters only the scale of the exponential tail; this gives the loss $\mathsf{w} + \mathsf{b}/4$ beside 2w. The half of the diagonal is a martingale plus a centered random drift, and both have variance $O(d_I)$, with no power of R_V. At each dyadic depth, the lengths d_I add up to t_N, so that the squared errors, before the normalization, add up without a loss in the number of queries. The fixed Gram event $\|\mathcal{T}^{\mathsf{T}}\mathcal{T}/n - I\| \le (16 D_*^2)^{-1}$, together with its cross block, fails with probability whose exponent loses $\mathsf{b}/2$ (Lemma 16.1). Given the scalar paths and the spins, Z is a centered Gaussian variable, since the scalar recursion and the link never read the diagonal entries or the unread block. Its conditional variance is at most $C_\beta n^{-1}\{h + \tau^2 + p(t_N)^2\}$ when $\|c - s\| \le \tau$, where $c_j = n^{-1}\langle u_j, x\rangle$ and s_j is the population value of $n^{-1}\langle \mathcal{T}_j, x\rangle$ (Lemma 17.2). The cruder bound $1/(2n)$ would again lose 2z.

Collecting these estimates, we obtain (2.12) with $\mathbb{P}(E_B^c) \le e^{-n^{1-\mathsf{s_b}}}$ whenever

$$\max\{2\mathsf{w}, \mathsf{w} + \mathsf{b}/4, \mathsf{b}/2, 2\mathsf{z} - \mathsf{a}, 2\mathsf{z} - 2\mathsf{w}, 2\mathsf{z} - \mathsf{b}\} < \mathsf{s_b},$$
$$\mathsf{b}/4 < \mathsf{a} < 1 - \mathsf{s_b}, \qquad \mathsf{k_b} < \mathsf{z} < \min(\mathsf{a}, \mathsf{b}/2, 2\mathsf{w}).$$

This is the second place where a sharp estimate is needed. The failure bound holds for the proposal law averaged over the disorder; no bound is asserted for each fixed disorder.

Smoothing, coverage, and the connecting sequence. A bound on $\mathrm{KL}(Q_F \| \Pi_n)$ does not imply that the proposal reaches a given set of appreciable Gibbs mass. To reverse the comparison, we follow [85, Section 5]. Conditionally on the spin x, the disorder under Π_n has density proportional to $e^{-\|g\|_F^2/2 + \beta H_g^c(x) - F^c(g)}$, and since F^c is convex and $H_g^c(x)$ is linear in g, this law is 1-strongly log-concave. The transport inequality of Bogachev, Kolesnikov, and Medvedev [24, Corollary 3.10] bounds the squared Wasserstein distance from it by twice the relative entropy (Lemma 18.1). We smooth the disorder by $g \mapsto \sqrt{1-\delta}\, g + \sqrt{\delta}\, z$, with $\delta = n^{-\mathsf{D}}$ and an independent standard Gaussian matrix z. As in [70, Lemma 11], the transport bound turns the forward entropy into a reverse entropy for the smoothed laws, at the cost of a factor δ^{-1}. Let $g_\star, g_\circ$ be independent with law γ_n, let $g_{\mathrm{base}} = \sqrt{1-\delta}\, g_\star + \sqrt{\delta}\, g_\circ$, and let $\nu^{\mathrm{sm}}_{g_{\mathrm{base}}}$ be the law of the spin of the fine proposal run on the matrix $\sqrt{1-\delta}\, g_{\mathrm{base}} + \sqrt{\delta}\, z$, with fresh auxiliary inputs. Then

$$\mathbb{E}\, \mathrm{KL}\big(\pi_{g_\star} \| \nu^{\mathrm{sm}}_{g_{\mathrm{base}}}\big) \le \frac{\beta^2 (n-1)}{2} \delta + \delta^{-1}\, \mathrm{KL}(Q_F \| \Pi_n)$$

(Lemma 18.2). Both terms are of order $n^{1-\mathsf{D}}$ and $n^{1-\mathsf{k_f}+\mathsf{D}}$, and the relative entropy is at most $n^{1-\mathsf{u_E}}$ with probability tending to one when $\mathsf{u_E} < \min(\mathsf{D}, \mathsf{k_f} - \mathsf{D})$. On this event, using the inequality for the binary relative entropy, we find that, simultaneously for all $S \subseteq \Sigma_n$,

$$\pi_{g_\star}(S) \geq n^{-\mathsf{d}} \quad \Longrightarrow \quad \nu^{\mathrm{sm}}_{g_{\mathrm{base}}}(S) \geq \exp\bigl\{-n^{\mathsf{d}}\bigl(n^{1-\mathsf{u_E}} + \log 2\bigr)\bigr\}$$

(Lemma 18.3). The set S may depend on $g_\star$. Two independent fine spins with law $\nu^{\mathrm{sm}}_{g_{\mathrm{base}}}$ fall on opposite sides of every cut with both masses at least $n^{-\mathsf{d}}$ with probability $\varpi_n \geq e^{-2n^{1-\mathsf{e}}}$ when $\mathsf{e} + \mathsf{d} < \mathsf{u_E}$.

Suppose now that $g_\star \in \mathcal{W}_n$, and let $S_{g_\star}$ be the first cut of small flow in a fixed order. The following argument is that of [85, Section 6]. If the two fine spins lie on opposite sides of $S_{g_\star}$, we join them by a sequence of configurations with short steps. Every such sequence comes close to the boundary of the cut. An enlargement of the boundary has small Gibbs mass, and the likelihood bound keeps the intermediate configurations out of it. We now describe the three ingredients.

The excluded set. Since $\|W(g_\star)\|_{\mathrm{op}} \leq 12$, every one-spin energy change is at most $24\sqrt{n}$, and every flip rate is at least $e^{-24\beta\sqrt{n}}/2$. Then the vertex boundary of $S_{g_\star}$ has Gibbs mass at most $4\phi_n e^{24\beta\sqrt{n}}$. We enlarge it by the Hamming radius $R_n = \lceil n^{1-\mathsf{r}} \rceil$, and to compare Gibbs weights on the enlargement, we also exclude configurations x for which

$$\sum_i \bigl(\mathbf{1}_A(i) - \mathbf{1}_B(i)\bigr)\bigl(W(g_\star)x\bigr)_i \geq n^{1-\mathsf{u_F}}$$

for some $A, B \subseteq [n]$ with $|A|, |B| \leq R_n$. The expected Gibbs mass of the resulting set $\mathcal{E}_n(g_\star)$ is at most $e^{-n^{1-\mathsf{q}}/2}$ when $0 < \mathsf{q} < \mathsf{u_F} < \mathsf{r} < 1$, $2\mathsf{u_F} - \mathsf{r} < \mathsf{q}$, and $\mathsf{q} < 1/2$ (Lemma 18.4). The bound on the local fields uses the transport inequality for the conditional disorder law given the spin. The cut is selected from $g_\star$ alone, and we also restrict to the event $F^c(g_\star) \leq nP_\beta + n^{1-\mathsf{z_P}}$, which fails with probability at most $e^{-c_\beta n^{1-2\mathsf{z_P}}}$ (Lemma 13.3).

The sequence. With independent standard Gaussian matrices g_1, g_2, we set

$$g_\theta = \sqrt{1-\delta}\, g_{\mathrm{base}} + \sqrt{\delta}\,(\cos\theta\, g_1 + \sin\theta\, g_2), \qquad 0 \leq \theta \leq \pi/2,$$

and we rotate two independent copies of the other Gaussian inputs of the proposal in the same way. For each fixed θ, the matrix g_θ has law γ_n, and the rotated inputs have their original law. We run the coarse proposal at the $\ell_n + 1$ points of a grid of spacing of order $e^{-n^{\mathsf{g_{rot}}}}$, and the fine proposal only at the two endpoints $\theta = 0$ and $\theta = \pi/2$. Given $(g_\star, g_\circ)$, the two fine spins are independent with law $\nu^{\mathrm{sm}}_{g_{\mathrm{base}}}$. The sequence consists of the first fine spin, the coarse spins in the order of the grid, and the second fine spin. On an event whose complement has probability at most $e^{-n^{1-\mathsf{s_{seq}}}}$, the following holds (Proposition 18.8):

- every coarse spin avoids $\mathcal{E}_n(g_\star)$: by the regression $g_\theta = (1-\delta)g_\star + \sqrt{\delta(2-\delta)}\, Z_\theta$, with Z_θ independent of $g_\star$, the conditional expectation of $e^{\beta H^c_{g_\theta}(x)}$ given $g_\star$ is at most $e^{O_\beta(n\delta)} e^{\beta H^c_{g_\star}(x)}$ when $\|W(g_\star)\|_{\mathrm{op}} \le 12$, and (2.11) bounds the probability of hitting the excluded set by $e^{-n^{1-\mathsf{q}}/4}$ when $\mathsf{q} < \min(\mathsf{k_b}, \mathsf{z_P}, \mathsf{D})$ (Lemma 18.5);
- consecutive coarse magnetizations differ by at most $e^{-n^{\mathsf{g_{rot}}}/2}$ when $\mathsf{a} < \mathsf{g_{rot}}$, because the recursion amplifies differences of its inputs by at most $e^{C_\beta n^{\mathsf{a}} \log n}$ on the Gram events at both ends of a grid step (Lemma 18.6);
- every spin, coarse or fine, is within squared Euclidean distance $R_n/4$ of its coarse magnetization when $\mathsf{r} < \mathsf{b}/2$. The fine and coarse runs share their paths up to V_b; the expected squared discrepancy per coordinate is $p(V_\mathsf{b}) \le C_\beta n^{-\mathsf{b}/2}$. The expected total squared Euclidean distance is $np(V_\mathsf{b}) = o(R_n)$; independence across rows gives the tail in Lemma 18.7.

On this event, we have that every step of the sequence has Hamming length at most R_n, and every step contains a coarse spin. The union over the grid costs a factor $e^{O(n^{\mathsf{g_{rot}}})}$, which is absorbed when $\mathsf{g_{rot}} < 1 - \mathsf{s_b}$, and the comparison of the failures requires $\mathsf{r} < \mathsf{s_b}$, $\mathsf{q} < \mathsf{s_b}$, and $\mathsf{s_b} < \mathsf{s_{seq}}$.

The fine proposal enters only through its relative entropy, which has no exceptional event, and through the two endpoint spins, which are joined to the grid by pinning. Every event in the union over the grid concerns the coarse proposal, so that the sequence needs no Gram event of the fine proposal and only the failure bound $e^{-n^{1-\mathsf{s_b}}}$ of the coarse one.

The contradiction. If the fine spins lie on opposite sides of $S_{g_\star}$, some step of the sequence crosses the cut. A shortest path across this step meets the boundary, so that the coarse endpoint of the step lies within distance R_n of the boundary, in $\mathcal{E}_n(g_\star)$. On the event of the sequence, this is impossible. Given an outer pair $(g_\star, g_\circ)$ in the event of the coverage bound, with $g_\star$ in $\mathcal{W}_n$ and in the pressure event, we conclude that ϖ_n is at most the conditional probability that the sequence fails. Integrating over the outer pair, and paying the coverage and pressure exceptions once, we find that

$$\gamma_n(\mathcal{W}_n) \le C_\beta\big(n^{\mathsf{u_E}-\mathsf{D}} + n^{\mathsf{u_E}-\mathsf{k_f}+\mathsf{D}}\big) + e^{-c_\beta n^{1-2\mathsf{z_P}}} + e^{2n^{1-\mathsf{e}} - n^{1-\mathsf{s_{seq}}}} \tag{2.14}$$

(Proposition 18.9), which tends to zero when $\mathsf{s_{seq}} < \mathsf{e}$.

The parameters. The exponents of Table 2.3 are affine functions of $\varepsilon_\omega = (1/4 - \omega)/20 \in (0, 1/80)$, and they satisfy every condition stated above, with strict inequalities, for every $\omega \in (0, 1/4)$ (Proposition 19.2). The complete list of conditions is given in Section 19.5. Four chains among them limit the saving:

$$\mathsf{k_f} < \min(\mathsf{a}, \mathsf{f}, 1-\mathsf{a}) \le \frac{1}{2}, \qquad \mathsf{e} + \mathsf{d} < \mathsf{u_E} < \min(\mathsf{D}, \mathsf{k_f} - \mathsf{D}) \le \frac{\mathsf{k_f}}{2},$$
$$\mathsf{q} < \mathsf{s_b} < \mathsf{s_{seq}} < \mathsf{e}, \qquad \omega < \mathsf{q}.$$

They come from the forward entropy, from smoothing and coverage, from the connecting sequence, and from (2.8). Together they give

$$\omega < \mathsf{q} < \mathsf{s}_{\mathrm{b}} < \mathsf{s}_{\mathrm{seq}} < \mathsf{e} < \mathsf{u}_{\mathrm{E}} < \frac{\mathsf{k}_{\mathrm{f}}}{2} \le \frac{1}{4}.$$

The conditions of coarse domination fit inside the third chain: the family has $\mathsf{q} < \mathsf{k}_{\mathrm{b}} < \mathsf{z} < \mathsf{b}/2 < \mathsf{s}_{\mathrm{b}}$, with consecutive gaps ε_ω, ε_ω, ε_ω, and $2\varepsilon_\omega$, and a loss $2\mathsf{z}$ in place of $2\mathsf{z} - \mathsf{a}$ would not fit. For $\omega = 1/5$, we have $\varepsilon_\omega = 1/400$, the chain reads

$$\frac{1}{5} < \frac{9}{40} < \frac{19}{80} < \frac{6}{25} < \frac{97}{400} < \frac{49}{200} < \frac{99}{400},$$

and (2.14) becomes $C_\beta n^{-1/400} + e^{-c_\beta n^{2/5}} + \exp\{2n^{303/400} - n^{19/25}\}$.

Both arguments stop at the exponent $1/4$. The proof of Theorem 1.1 needs $n\rho^4/\log n \to \infty$ for the overlap scale ρ, and the proof of Theorem 1.3 needs the chain of inequalities $\omega < \mathsf{q} < \mathsf{s}_{\mathrm{b}} < \mathsf{s}_{\mathrm{seq}} < \mathsf{e} < \mathsf{u}_{\mathrm{E}} < \mathsf{k}_{\mathrm{f}}/2 \le 1/4$ of Section 19.5. We do not know whether this coincidence has a common explanation.

2.8 Fluctuations and the mean bias

Theorem 1.6 is OpenAI's theorem [81, Theorem 1.1]. We give no new proof of it; our contribution concerning it and Corollary 1.7 is described in Section 22. In [81], the theorem is reduced to a two-sided bound on a positive logarithmic moment. Let $F_n^c = F_n + \beta G$ be the completed log partition function, and let $K_n(t) = \log \mathbb{E} \exp\{t(F_n^c - \mathbb{E}F_n^c)\}$. For fixed $m \in (0, 1/300)$ and $s_n = n^{-1/6+5m}$, it is proved in [81, Proposition 7.5] that $cn^{20m} \le K_n(s_n) \le Cn^{100m}$ for all sufficiently large n (Proposition 20.1); Section 20.2 summarizes this proof with locators. OpenAI bounds $K_n(s_n)$ by comparison with finite Gaussian hierarchies along overlap quantiles. The error of such a comparison is small when its minimizers satisfy a two-sided linear bound. This bound follows from a one-site cavity estimate [81, Proposition 4.1], from the asymptotic linearity of the rescaled overlap quantiles of the minimizers [81, Theorem 5.1], and from finite propagation through a bounded number of successive comparisons [81, Proposition 6.6]. The lower bound uses a chord inequality at $\varepsilon = n^{-1/3}$, and the upper bound an iteration in ε with a low-field perturbation [81, Sections 7.1–7.4].

A bound on K_n at one tilt does not control $\mathrm{Var}(F_n)$ by itself, since a rare deviation can dominate an exponential moment or a variance. The additional input is that the sum of F_n^c and a random variable with the exponential law of rate one, independent of (J, G), has a log-concave density (Lemma 20.2). This observation is due to Aronow and Lopatto [8, Proposition 5.1], and [81, Lemma 7.6] states it in the present setting. For a log-concave law with standard deviation σ, the centered logarithmic moment is bounded at tilts below c/σ and grows at least linearly in $t\sigma$ at larger tilts t, and every interval of length $2r$ has

probability at most Cr/σ (Lemma 20.3). We deduce that $\mathrm{Var}(F_n) = n^{1/3+o(1)}$ and, using the bound on intervals, the lower bound on $|F_n - \mathbb{E}F_n|$ in (1.6) (Sections 20.3 and 20.4, following [81, Sections 7.5 and 7.6]).

Corollary 1.7 is the lower bound $n^{-5/6-\varepsilon}$ on the mean bias. It follows from the variance–bias inequality $\Delta_{n,\beta} \geq c_\beta((\mathrm{Var}(F_n) + \beta^2/2)/n)^{5/4}$ of Aronow and Lopatto [8, Lemma 7.4 and Remark 7.14] (Proposition 21.1). Inserting $\mathrm{Var}(F_n) \geq n^{1/3-4\varepsilon/5}$ from Theorem 1.6, we find that $\Delta_{n,\beta} \geq c_\beta n^{-5/6-\varepsilon}$ for all sufficiently large n. Together with the upper bound $\Delta_{n,\beta} \leq C_b n^{-7/10}$ of Theorem 4.11, which determines the exponent of Theorem 1.1, this shows that if $\Delta_{n,\beta} = n^{-a+o(1)}$ for some a, then $7/10 \leq a \leq 5/6$. Both deductions are already recorded in [6, Section 1.6].

PART I

Preliminaries

SECTION 3

The model and the dynamics

This section fixes the model, the dynamics, and the conventions used in the rest of the paper. Section 3.1 defines the Sherrington–Kirkpatrick model, its Gibbs measure, the overlap, the log partition function F_n, the limiting pressure P_β, and the mean bias $\Delta_{n,\beta}$. Section 3.2 records exact identities for the Hamiltonian completed by an independent Gaussian variable (Lemma 3.1). They are used in Section 4 to translate imported results into our normalization, in Section 8 for fractional moments of constrained partition functions, in Part IV for the Gaussian-query proposal, and in Sections 20 and 21. Section 3.3 defines heat-bath dynamics in discrete and continuous time and proves the elementary facts collected in Lemma 3.2: reversibility, Poissonization, and the contraction of total variation. These facts are used in Section 12 to prove Theorem 1.1 and Corollary 1.2 and in Part IV to prove Theorem 1.3 and Corollary 1.4. Section 3.4 defines the realized-start probability in which Theorems 1.1 and 1.3 are formulated, and Lemma 3.4 relates statements about this probability to statements about the Gibbs mass of the corresponding initial states. Section 3.5 fixes the asymptotic conventions, and Section 3.6 compares our normalization with those of the works from which we import results.

The model, the heat-bath dynamics, and the realized-start formulation are those of [77, Section 1]. The facts proved in this section are standard. We include short proofs so that the conventions can be checked directly, including the clock and the order in which the disorder and the initial state are sampled.

3.1 The Sherrington–Kirkpatrick model

We consider the zero-field Gaussian Sherrington–Kirkpatrick model, introduced in [100]. For an integer $n \geq 2$, let $\Sigma_n = \{-1, 1\}^n$. Its elements, usually denoted x, y, z, or v, are called configurations. The disorder is a family $J = (J_{ij})_{1 \leq i < j \leq n}$ of independent standard Gaussian random variables, and we set $J_{ji} = J_{ij}$ and $J_{ii} = 0$. We write $\mathbb{P}_J$ and $\mathbb{E}_J$ for probability and expectation over J. Unsubscripted $\mathbb{P}$, $\mathbb{E}$, and Var refer to all random variables under consideration, except those declared fixed. In Section 3.2 they average over (J, G), whereas in Lemma 3.2 the disorder is held fixed and probabilities average over the chain and the independent clock. For $\beta > 0$ and $x \in \Sigma_n$, we define the Hamiltonian,

the partition function, and the log partition function by

$$H_n(x) = \frac{1}{\sqrt{n}} \sum_{1 \le i < j \le n} J_{ij} x_i x_j, \qquad Z_n = \sum_{x \in \Sigma_n} e^{\beta H_n(x)}, \qquad F_n = \log Z_n. \tag{3.1}$$

When the dependence on the disorder or on β matters, we write H_n^J and $Z_n = Z_{n,\beta}^J$. The Gibbs measure is the probability measure on Σ_n given by

$$\pi(x) = \pi^J(x) = \pi_{n,\beta}^J(x) = \frac{e^{\beta H_n^J(x)}}{Z_{n,\beta}^J}, \qquad x \in \Sigma_n. \tag{3.2}$$

For a function f on Σ_n and a set $A \subseteq \Sigma_n$, we write $\langle f \rangle = \pi(f) = \sum_x f(x)\pi(x)$ and $\pi(A) = \sum_{x \in A} \pi(x)$.

Three features of (3.1) are used throughout. First, the inverse temperature multiplies H_n and is not part of it. Second, Z_n is a sum with respect to counting measure on Σ_n, with no factor 2^{-n}. Third, F_n is the logarithm of the partition function, not a free energy per spin. In physical terms, the energy is $-H_n$ and the free energy is $-F_n/\beta$; since β is fixed, the factor $-1/\beta$ does not affect any exponent in this paper.

The overlap of two configurations $x, y \in \Sigma_n$ is

$$R(x,y) = \frac{1}{n} \sum_{i=1}^n x_i y_i \in [-1,1].$$

We have $R(x,x) = 1$, $R(-x,y) = -R(x,y)$, and $H_n(-x) = H_n(x)$, the last because H_n is quadratic in the spins. It follows that the Gibbs measure is invariant under the global spin reversal $x \mapsto -x$. The Hamiltonian is a centered Gaussian process indexed by Σ_n, with covariance

$$\mathbb{E} H_n(x) H_n(y) = \frac{n R(x,y)^2 - 1}{2}, \qquad x, y \in \Sigma_n. \tag{3.3}$$

To see this, set $a_i = x_i y_i$. Using that $\mathbb{E} J_{ij} J_{kl} = \mathbf{1}\{(i,j) = (k,l)\}$ for $i < j$ and $k < l$, and that $a_i^2 = 1$, we find

$$\mathbb{E} H_n(x) H_n(y) = \frac{1}{n} \sum_{i<j} a_i a_j = \frac{1}{2n} \Bigl(\Bigl(\sum_{i=1}^n a_i \Bigr)^2 - \sum_{i=1}^n a_i^2 \Bigr) = \frac{n^2 R(x,y)^2 - n}{2n}.$$

We have $\operatorname{Var} H_n(x) = (n-1)/2$ for all x. The constant $-1/2$ in (3.3) reflects the absence of diagonal terms in H_n; it is removed in Section 3.2.

We write

$$p_n(\beta) = \frac{1}{n} \mathbb{E} F_n$$

for the expected pressure. It is finite, since $\beta H_n(x_0) \le F_n \le n \log 2 + \beta \max_x H_n(x)$ for any fixed $x_0 \in \Sigma_n$ and each $H_n(x)$ is Gaussian. In [60, Theorems 1–2, (17)–(18)], Guerra and Toninelli proved that the sequence $n\, p_n(\beta)$ is superadditive in n and that the limit

$$P_\beta = \lim_{n \to \infty} p_n(\beta)$$

exists for every $\beta > 0$ and equals the supremum of the sequence. We call

$$\Delta_{n,\beta} = P_\beta - p_n(\beta) \tag{3.4}$$

the mean bias. We have $\Delta_{n,\beta} \geq 0$ for all $n \geq 2$, because P_β is the supremum of the sequence $p_n(\beta)$; this is also a consequence of Guerra's bound [59, Theorem 3]. The identification of P_β with the minimum of the Parisi functional, due to Guerra [59] and Talagrand [103], is stated in Section 4 as Theorem 4.4, where the Parisi functional is also defined.

3.2 The completed Hamiltonian

By (3.3), the covariance of H_n differs from $nR(x,y)^2/2$ by a constant. Several of the results we import, and the Gaussian interpolation of Section 8, are formulated for a Hamiltonian whose covariance is exactly $nR(x,y)^2/2$. Such a Hamiltonian is obtained by adding a single Gaussian variable, the same for every configuration. Let G be a centered Gaussian random variable with variance $1/2$, independent of J, and set

$$H_n^c(x) = H_n(x) + G, \qquad Z_n^c = \sum_{x \in \Sigma_n} e^{\beta H_n^c(x)} = e^{\beta G} Z_n, \qquad F_n^c = \log Z_n^c = F_n + \beta G. \tag{3.5}$$

We call H_n^c the completed Hamiltonian and F_n^c the completed log partition function. In [6, 8, 77], the variable G arises from the diagonal of a symmetric Gaussian matrix. Let $J'_{11}, \dots, J'_{nn}$ be independent standard Gaussian variables, independent of J, and let W be the symmetric $n \times n$ matrix with $W_{ij} = J_{ij}/\sqrt{n}$ for $i \neq j$ and $W_{ii} = \sqrt{2/n}\, J'_{ii}$. Then

$$\frac{1}{2} \sum_{i,j=1}^{n} W_{ij} x_i x_j = H_n(x) + \frac{1}{\sqrt{2n}} \sum_{i=1}^{n} J'_{ii},$$

and the last term has the law of G.

Lemma 3.1 (Exact identities for the completed Hamiltonian)**.** *Fix $n \geq 2$ and $\beta > 0$. Then the following statements hold.*

(a) For all $x, y \in \Sigma_n$, we have $\mathbb{E} H_n^c(x) H_n^c(y) = nR(x,y)^2/2$.

(b) For every realization of (J, G) and every $x \in \Sigma_n$, we have $e^{\beta H_n^c(x)}/Z_n^c = \pi^J(x)$.

(c) We have $\mathbb{E} F_n^c = \mathbb{E} F_n$ and $\operatorname{Var} F_n^c = \operatorname{Var} F_n + \beta^2/2$. Further, for every $t \in \mathbb{R}$, both sides of the following identity are finite, and

$$\log \mathbb{E} \exp\{t(F_n^c - \mathbb{E} F_n^c)\} = \log \mathbb{E} \exp\{t(F_n - \mathbb{E} F_n)\} + \frac{\beta^2 t^2}{4}. \tag{3.6}$$

(d) *Let $\ell \geq 1$ be an integer, and let $A \subseteq \Sigma_n^\ell$ be a nonempty set of ℓ-tuples of configurations. Define*

$$Z_{\ell,n}(A) = \sum_{(x^1,\dots,x^\ell)\in A} \exp\Big(\beta \sum_{a=1}^{\ell} H_n(x^a)\Big),$$

and define $Z^c_{\ell,n}(A)$ in the same way, with H^c_n in place of H_n. Then $Z^c_{\ell,n}(A) = e^{\beta\ell G} Z_{\ell,n}(A)$ and $\mathbb{E}\log Z^c_{\ell,n}(A) = \mathbb{E}\log Z_{\ell,n}(A)$. For every $s \in \mathbb{R}$, both sides of the following identity are finite, and

$$\mathbb{E} Z^c_{\ell,n}(A)^s = \exp\Big(\frac{\beta^2\ell^2 s^2}{4}\Big)\, \mathbb{E} Z_{\ell,n}(A)^s. \tag{3.7}$$

Equivalently, for every $s \neq 0$,

$$\frac{1}{ns}\log \mathbb{E} Z^c_{\ell,n}(A)^s = \frac{1}{ns}\log \mathbb{E} Z_{\ell,n}(A)^s + \frac{\beta^2\ell^2 s}{4n}.$$

Proof. Part (a) follows from (3.3), since G is independent of J and $\mathbb{E}G^2 = 1/2$. For part (b), the factor $e^{\beta G}$ appears in both $e^{\beta H^c_n(x)}$ and Z^c_n.

For part (c), recall that F_n is a function of J and that G is independent of J. Using $\mathbb{E}G = 0$ and $\operatorname{Var}(\beta G) = \beta^2/2$, we have $\mathbb{E}F^c_n = \mathbb{E}F_n$ and $\operatorname{Var} F^c_n = \operatorname{Var} F_n + \beta^2/2$. Next, we check that all exponential moments of F_n are finite. Fix $x_0 \in \Sigma_n$. Since $e^{\beta H_n(x_0)} \leq Z_n \leq 2^n \max_x e^{\beta H_n(x)}$, we have

$$e^{tF_n} \leq 2^{nt} \sum_{x\in\Sigma_n} e^{t\beta H_n(x)} \quad (t \geq 0), \qquad e^{tF_n} \leq e^{t\beta H_n(x_0)} \quad (t < 0).$$

As each $H_n(x)$ is Gaussian, this implies that $\mathbb{E}e^{tF_n} < \infty$ for all $t \in \mathbb{R}$. By the independence of F_n and G and the identity $\mathbb{E}e^{t\beta G} = e^{\beta^2 t^2/4}$, we have

$$\mathbb{E}\exp\{t(F^c_n - \mathbb{E}F^c_n)\} = \mathbb{E}\exp\{t(F_n - \mathbb{E}F_n)\}\, \mathbb{E}e^{t\beta G} = \mathbb{E}\exp\{t(F_n - \mathbb{E}F_n)\}\, e^{\beta^2 t^2/4}.$$

Taking logarithms gives (3.6).

For part (d), note that every term in the sum defining $Z^c_{\ell,n}(A)$ contains the factor $\exp(\beta\sum_{a=1}^{\ell} G) = e^{\beta\ell G}$, which proves $Z^c_{\ell,n}(A) = e^{\beta\ell G} Z_{\ell,n}(A)$. Fix a tuple $(x^1_0,\dots,x^\ell_0) \in A$. The bounds of part (c), with $\sum_a H_n(x^a_0)$ in place of $H_n(x_0)$ and $|A|$ in place of 2^n, show that $\log Z_{\ell,n}(A)$ is integrable and that $\mathbb{E}Z_{\ell,n}(A)^s < \infty$ for every $s \in \mathbb{R}$, since $\sum_a H_n(x^a)$ is Gaussian for each tuple $(x^1,\dots,x^\ell)$. Taking logarithms and expectations in $Z^c_{\ell,n}(A) = e^{\beta\ell G} Z_{\ell,n}(A)$ and using $\mathbb{E}G = 0$, we find $\mathbb{E}\log Z^c_{\ell,n}(A) = \mathbb{E}\log Z_{\ell,n}(A)$. The variable $\beta\ell s G$ is centered Gaussian with variance $\beta^2\ell^2 s^2/2$, and it is independent of $Z_{\ell,n}(A)$. We conclude that

$$\mathbb{E}Z^c_{\ell,n}(A)^s = \mathbb{E}e^{\beta\ell s G}\, \mathbb{E}Z_{\ell,n}(A)^s = e^{\beta^2\ell^2 s^2/4}\, \mathbb{E}Z_{\ell,n}(A)^s,$$

which is (3.7). Dividing its logarithm by ns gives the last display of the lemma. □

The constrained partition functions of Sections 4, 8, and 9, in which the overlaps $R(x^a, x^b)$ of the ℓ-tuple are prescribed, are of the form $Z_{\ell,n}(A)$. By part (b) of Lemma 3.1 and the identity $Z^c_{\ell,n}(A) = e^{\beta\ell G} Z_{\ell,n}(A)$, Gibbs measures, constrained Gibbs probabilities $Z_{\ell,n}(A)/Z_n^\ell$, and expected logarithms are the same for the two Hamiltonians. Fractional moments are not: the correction $\beta^2\ell^2 s/(4n)$ in the last display of the lemma is quadratic in ℓ, because the same variable G enters every replica, and it is kept explicitly in Section 8. The identity (3.6) is the form in which the completion enters Section 20, where the logarithmic-moment bounds of [81] are stated for F_n^c.

3.3 Heat-bath dynamics

We use single-spin heat-bath dynamics, in the tradition of the stochastic Ising dynamics introduced by Glauber [58]. The clock and resampling conventions below are those of [77, Section 1, pp. 2–3]. For $x \in \Sigma_n$, $i \in \{1, \dots, n\}$, and $\sigma \in \{-1, 1\}$, let $x^{i,\sigma}$ be the configuration obtained from x by setting its ith spin equal to σ. The local field at site i and the heat-bath probabilities are

$$h_i(x) = h_i^J(x) = \frac{1}{\sqrt{n}} \sum_{j \neq i} J_{ij} x_j, \qquad p_i(\sigma \mid x) = p_i^J(\sigma \mid x) = \frac{e^{\beta\sigma h_i(x)}}{2\cosh(\beta h_i(x))}. \tag{3.8}$$

The field $h_i(x)$ does not depend on x_i. Since $J_{ji} = J_{ij}$, the terms of H_n that contain x_i add up to $x_i h_i(x)$, and we have

$$H_n(x) = x_i h_i(x) + H_n^{(i)}(x), \qquad H_n^{(i)}(x) = \frac{1}{\sqrt{n}} \sum_{\substack{j<k \\ j,k \neq i}} J_{jk} x_j x_k, \tag{3.9}$$

where $H_n^{(i)}(x)$ does not depend on x_i. It follows that $p_i(\sigma \mid x) = \pi(x^{i,\sigma})/(\pi(x^{i,1}) + \pi(x^{i,-1}))$; that is, $p_i(\cdot \mid x)$ is the conditional law of the ith spin under π, given the other spins.

The discrete heat-bath kernel is

$$K(x, y) = K^J(x, y) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\sigma \in \{-1,1\}} p_i(\sigma \mid x)\, \mathbf{1}\{y = x^{i,\sigma}\}, \qquad x, y \in \Sigma_n. \tag{3.10}$$

In one step, a site i is chosen uniformly at random and the ith spin is resampled from $p_i(\cdot \mid x)$. The new value may equal the old one, and in this case the update attempt leaves the configuration unchanged. Explicitly, for all $x \in \Sigma_n$ and all sites i,

$$K(x, x^{i,-x_i}) = \frac{1}{n}\, p_i(-x_i \mid x), \qquad K(x, x) = \frac{1}{n} \sum_{i=1}^{n} p_i(x_i \mid x), \tag{3.11}$$

and $K(x,y) = 0$ if x and y differ in more than one coordinate. The discrete heat-bath chain $(X_k)_{k\geq 0}$ is the Markov chain on Σ_n with transition matrix K.

In continuous time, we define the generator and the transition probabilities by

$$\mathcal{L} = \mathcal{L}^J = n(K - I), \qquad P_t = P_t^J = e^{t\mathcal{L}}, \quad t \geq 0, \tag{3.12}$$

and we call the Markov process $(X_t)_{t\geq 0}$ with transition probabilities P_t the continuous heat-bath process. In this process, each site carries an independent Poisson clock of rate one, and when the clock at site i rings, the ith spin of the current configuration x is resampled from $p_i(\cdot \mid x)$. Indeed, the generator of this description acts on functions $f : \Sigma_n \to \mathbb{R}$ by

$$\mathcal{L}f(x) = \sum_{i=1}^{n} \sum_{\sigma\in\{-1,1\}} p_i(\sigma \mid x)\big(f(x^{i,\sigma}) - f(x)\big) = n\big(Kf(x) - f(x)\big),$$

where the second equality follows from (3.10) and $\sum_\sigma p_i(\sigma \mid x) = 1$. The total rate of update attempts is n, so that one unit of continuous time corresponds to n update attempts on average. In both time conventions, an attempt that leaves the configuration unchanged is counted. We use the letter t for continuous time and k for the number of discrete steps, so that the notation (X_t) or (X_k) indicates which process is meant.

For probability measures ν, ζ on Σ_n, the total variation distance is

$$\|\nu - \zeta\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x\in\Sigma_n} |\nu(x) - \zeta(x)| = \max_{A\subseteq\Sigma_n} |\nu(A) - \zeta(A)|,$$

where the second equality is [68, Proposition 4.2]. For $x \in \Sigma_n$, $t \geq 0$, and integers $k \geq 0$, we define

$$d^J(x,t) = \|P_t^J(x,\cdot) - \pi^J\|_{\mathrm{TV}}, \qquad d^J_{\mathrm{disc}}(x,k) = \|(K^J)^k(x,\cdot) - \pi^J\|_{\mathrm{TV}}. \tag{3.13}$$

These are the distances from equilibrium of the laws of the dynamics started at the fixed configuration x. No supremum over x is taken. For the threshold $1/4$, we define the mixing times from x by

$$\begin{aligned} T^J(x) &= \inf\{t \geq 0 : d^J(x,t) \leq 1/4\}, \\ T^J_{\mathrm{disc}}(x) &= \min\{k \geq 0 : d^J_{\mathrm{disc}}(x,k) \leq 1/4\}. \end{aligned} \tag{3.14}$$

Both times are finite. Indeed, every single-site resampling probability is positive, so $K^n(x,y) > 0$ for all $x, y \in \Sigma_n$: one may update each site once to its value in y. The minimum of these finitely many entries is positive. Writing K^n as a mixture with a positive uniform component and using contraction shows that $d^J_{\mathrm{disc}}(x,k) \to 0$; (3.15) then gives $d^J(x,t) \to 0$. By Lemma 3.2(d), the continuous sublevel set in (3.14) is closed and

upward closed. For every $t \geq 0$, we have $T^J(x) \leq t$ if and only if $d^J(x,t) \leq 1/4$, and $T^J(x) > t$ if and only if $d^J(x,t) > 1/4$. The corresponding equivalences hold in discrete time.

In the next lemma, we record classical facts about finite Markov chains for the kernels K and P_t; see [68, Proposition 1.20, Exercise 4.2, and Section 20.1].

Lemma 3.2 (Elementary properties of heat-bath dynamics)**.** *Fix $n \geq 2$, $\beta > 0$, and a realization of J. Then the following statements hold.*

(a) The kernel K is reversible with respect to π, that is, $\pi(x)K(x,y) = \pi(y)K(y,x)$ for all $x, y \in \Sigma_n$. In particular, $\pi K = \pi$.

(b) For every $t \geq 0$,

$$P_t = \sum_{k=0}^{\infty} e^{-nt} \frac{(nt)^k}{k!} K^k. \tag{3.15}$$

If $(X_k)_{k\geq 0}$ is the discrete heat-bath chain started at x and $(N_t)_{t\geq 0}$ is a Poisson process of rate n, independent of $(X_k)_{k\geq 0}$, then $\mathbb{P}(X_{N_t} = y) = P_t(x,y)$ for all $y \in \Sigma_n$ and $t \geq 0$.

(c) For every $t \geq 0$, the kernel P_t is reversible with respect to π, and $\pi P_t = \pi$. In particular, $\sum_x \pi(x)P_t(x,\cdot) = \pi$ and $\sum_x \pi(x)K^k(x,\cdot) = \pi$ for all $t \geq 0$ and $k \geq 0$.

(d) For every transition matrix Q on Σ_n and all probability measures ν, ζ on Σ_n, we have $\|\nu Q - \zeta Q\|_{\mathrm{TV}} \leq \|\nu - \zeta\|_{\mathrm{TV}}$. For every $x \in \Sigma_n$, the functions $k \mapsto d^J_{\mathrm{disc}}(x,k)$ and $t \mapsto d^J(x,t)$ are nonincreasing, and $t \mapsto d^J(x,t)$ is continuous.

Proof. For part (a), the identity is trivial if $x = y$, and both sides vanish by (3.11) if x and y differ in more than one coordinate. Suppose that $y = x^{i,-x_i}$ for some i. Then $h_i(y) = h_i(x)$ and $H_n^{(i)}(y) = H_n^{(i)}(x)$. Using (3.8), (3.9), and (3.11), we find

$$\pi(x)K(x,y) = \frac{e^{\beta H_n^{(i)}(x)}}{nZ_n} \cdot \frac{e^{\beta x_i h_i(x)} e^{\beta y_i h_i(x)}}{2\cosh(\beta h_i(x))}.$$

The right side is symmetric in x and y, which proves the identity. Summing the identity $\pi(x)K(x,y) = \pi(y)K(y,x)$ over x and using $\sum_x K(y,x) = 1$, we obtain $\pi K = \pi$.

For part (b), since the matrices ntK and ntI commute, we have $e^{tn(K-I)} = e^{-nt}e^{ntK}$, and expanding e^{ntK} in its power series gives (3.15). For the second statement, recall that N_t has the Poisson law with mean nt and is independent of $(X_k)_{k\geq 0}$. Conditioning on N_t, we find

$$\mathbb{P}(X_{N_t} = y) = \sum_{k=0}^{\infty} \mathbb{P}(N_t = k)\, K^k(x,y) = \sum_{k=0}^{\infty} e^{-nt} \frac{(nt)^k}{k!} K^k(x,y) = P_t(x,y).$$

For part (c), we first show by induction on k that $\pi(x)K^k(x,y) = \pi(y)K^k(y,x)$ for all $k \geq 0$. The case $k = 0$ is trivial. If the identity holds for k, then by part (a) we have

$$\pi(x)K^{k+1}(x,y) = \sum_z \pi(z)K(z,x)K^k(z,y) = \sum_z K(z,x)\, \pi(y)K^k(y,z) = \pi(y)K^{k+1}(y,x).$$

By (3.15), P_t is a convex combination of the matrices K^k, with weights that do not depend on x or y. This implies that $\pi(x)P_t(x,y) = \pi(y)P_t(y,x)$, and summing over x gives $\pi P_t = \pi$. Summing the identity for K^k over x in the same way, we find $\pi K^k = \pi$.

For part (d), for each y we have $|\nu Q(y) - \zeta Q(y)| \le \sum_x |\nu(x) - \zeta(x)| Q(x,y)$. Summing over y and using $\sum_y Q(x,y) = 1$ gives the first claim. Applying it with $Q = K$, $\nu = K^k(x,\cdot)$, and $\zeta = \pi$, and recalling that $\pi K = \pi$, we find

$$d^J_{\mathrm{disc}}(x, k+1) = \|K^k(x,\cdot)K - \pi K\|_{\mathrm{TV}} \le d^J_{\mathrm{disc}}(x,k).$$

Similarly, for $u \ge 0$ we have $P_{t+u} = P_t P_u$ and $\pi P_u = \pi$, and the first claim with $Q = P_u$ implies that $d^J(x, t+u) \le d^J(x,t)$. Finally, by (3.15), every entry of P_t is a continuous function of t. As a finite sum of absolute values of such functions, $d^J(x,t)$ is continuous in t. □

We can realize the continuous heat-bath process started at x as $(X_{N_t})_{t\ge 0}$, with $(X_k)_{k\ge 0}$ and $(N_t)_{t\ge 0}$ as in part (b) of Lemma 3.2. Indeed, the superposition of the n independent rate-one clocks is a Poisson process of rate n. At each of its points, the site that rings is uniform on $\{1,\dots,n\}$ and independent of the past; see [68, Section 20.1, (20.1)–(20.6)].

We use part (d) of Lemma 3.2 to turn a lower bound on the distance at one time into lower bounds at all earlier times. If $t \le t'$ and $d^J(x,t') > a$, then $d^J(x,t) > a$, and likewise in discrete time. This is the property used in the proof of Corollary 1.2, and parts (b) and (c) are used in the proof of Theorem 1.1. We note that part (c) concerns deterministic times. It says that a chain started from π has law π at every fixed time. It says nothing about the law of the chain at a random time that depends on the trajectory.

3.4 The realized-start experiment

In Theorems 1.1 and 1.3, the initial state of the dynamics is a Gibbs sample for the realized disorder, and it is held fixed when the distance from equilibrium is evaluated. We make this precise in the following definition.

Definition 3.3 (Realized-start probability)**.** Fix $n \ge 2$ and $\beta > 0$. Let $\mathbb{P}_{J,X_0\sim\pi^J}$ be the joint law of a pair (J, X_0) obtained by first sampling the disorder J and then sampling X_0 from π^J. That is, for every Borel set $B \subseteq \mathbb{R}^{n(n-1)/2}$ and every $x \in \Sigma_n$,

$$\mathbb{P}_{J,X_0\sim\pi^J}\{J \in B,\ X_0 = x\} = \mathbb{E}_J\big[\pi^J(x)\,\mathbf{1}\{J \in B\}\big].$$

For a deterministic time $t \ge 0$ and a threshold $a \in \mathbb{R}$, the realized-start probability is

$$\mathbb{P}_{J,X_0\sim\pi^J}\{d^J(X_0,t) > a\} = \mathbb{E}_J \sum_{x\in\Sigma_n} \pi^J(x)\,\mathbf{1}\{d^J(x,t) > a\} = \mathbb{E}_J\,\pi^J\{x \in \Sigma_n : d^J(x,t) > a\}, \tag{3.16}$$

and the discrete realized-start probability is defined in the same way, with $d^J_{\mathrm{disc}}(X_0,k)$ for a deterministic integer $k \ge 0$ in place of $d^J(X_0,t)$.

The first equality in (3.16) is the definition of the joint law, and the second is the definition of $\pi^J(\cdot)$ for sets. The quantity $d^J(X_0, t)$ is the distance from π^J of the law at time t of the dynamics started at X_0. Its randomness comes only from J and X_0, since the randomness of the dynamics is integrated in $P_t^J(X_0, \cdot)$.

We evaluate the distance before averaging over the initial state. If we averaged over the initial state first, the law at time t would be $\sum_x \pi^J(x) P_t^J(x, \cdot) = \pi^J$ by part (c) of Lemma 3.2, and its distance from π^J would be zero at every time. The realized-start probability measures instead how much Gibbs mass lies on initial states from which the dynamics is still far from equilibrium at time t. This is the formulation of [77, Section 1, following Theorem 1.1, p. 3]. It is not a worst-case statement: a lower bound on $\max_x d^J(x, t)$ gives no information about the Gibbs mass of the initial states from which the dynamics mixes slowly.

We note that all quantities in (3.16) are measurable. Since Σ_n is finite, the Gibbs weights $\pi^J(x)$ and the entries of K^J are continuous functions of J, and so are the entries of $(K^J)^k$ and of $P_t^J = e^{tn(K^J - I)}$. It follows that for all x, t, and k, the maps $J \mapsto d^J(x, t)$ and $J \mapsto d^J_{\mathrm{disc}}(x, k)$ are continuous, the sets $\{J : d^J(x, t) > a\}$ are open, and $J \mapsto \pi^J\{x : d^J(x, t) > a\}$ is Borel measurable. No measurable selection of initial states or of auxiliary objects is needed to define these probabilities.

Lemma 3.4 (Equivalent forms of a realized-start statement)**.** *Fix $\beta > 0$ and $a \in \mathbb{R}$, and let $(t_n)_{n \geq 2}$ be deterministic times. Set*

$$M_n(J) = \pi^J\{x \in \Sigma_n : d^J(x, t_n) > a\}.$$

Then the following three statements are equivalent.

(a) $\mathbb{P}_{J, X_0 \sim \pi^J}\{d^J(X_0, t_n) > a\} \to 1$ *as* $n \to \infty$.

(b) $\mathbb{E}_J M_n \to 1$ *as* $n \to \infty$.

(c) $M_n \to 1$ *in probability over the disorder; that is,* $\mathbb{P}_J\{M_n > 1 - \varepsilon\} \to 1$ *for every* $\varepsilon > 0$.

The same holds with $d^J_{\mathrm{disc}}(x, k_n)$, for deterministic integers $k_n \geq 0$, in place of $d^J(x, t_n)$. If these statements hold both in continuous and in discrete time, then $\mathbb{P}_{J, X_0 \sim \pi^J}\{d^J(X_0, t_n) > a,\ d^J_{\mathrm{disc}}(X_0, k_n) > a\} \to 1$, and the Gibbs mass of the set of x that satisfy both inequalities tends to one in probability over the disorder.

Proof. By (3.16), statements (a) and (b) are the same. Fix $\varepsilon \in (0, 1)$. Since $0 \leq 1 - M_n \leq 1$, by Markov's inequality we have

$$\mathbb{P}_J\{1 - M_n \geq \varepsilon\} \leq \frac{\mathbb{E}_J(1 - M_n)}{\varepsilon},$$

and splitting the expectation according to whether $1 - M_n < \varepsilon$, we also find

$$\mathbb{E}_J(1 - M_n) \leq \varepsilon + \mathbb{P}_J\{1 - M_n \geq \varepsilon\}.$$

By the first inequality, (b) implies (c). By the second inequality, (c) implies $\limsup_n \mathbb{E}_J(1 - M_n) \leq \varepsilon$ for every $\varepsilon \in (0,1)$, and hence (b) holds. The argument in discrete time is identical.

For the last statement, let M'_n be the Gibbs mass of $\{x : d^J_{\mathrm{disc}}(x, k_n) > a\}$, and let M''_n be the Gibbs mass of the set of x satisfying both inequalities. By a union bound for π^J, we have $1 - M''_n \leq (1 - M_n) + (1 - M'_n)$. Statement (b) holds for M_n and for M'_n, and this implies that $\mathbb{E}_J(1 - M''_n) \to 0$. By the definition of the joint law, $\mathbb{E}_J M''_n$ is the joint probability that X_0 satisfies both inequalities, and the equivalence of (b) and (c) for M''_n yields the second assertion. □

The complementary events satisfy the same equivalences. For deterministic t_n and a fixed a, the probability $\mathbb{P}_{J, X_0 \sim \pi^J}\{d^J(X_0, t_n) \leq a\}$ tends to one if and only if $\mathbb{P}_{J, X_0 \sim \pi^J}\{d^J(X_0, t_n) > a\}$ tends to zero. By (3.16), this is equivalent to $\mathbb{E}_J M_n \to 0$, and hence to $M_n \to 0$ in probability, since $0 \leq M_n \leq 1$. Equivalently, the Gibbs mass of $\{x : d^J(x, t_n) \leq a\}$ tends to one in probability over the disorder. This is the form used in Theorem 1.3. The same statements hold in discrete time.

By Lemma 3.4, each of the two conclusions of Theorem 1.1 is equivalent to the convergence to one, in probability over the disorder, of the Gibbs mass of the initial states that satisfy it, and the two conclusions also hold simultaneously. We use the lemma for each fixed threshold; it gives no uniformity in the threshold and no almost sure statement in the disorder.

3.5 Asymptotic conventions

The definitions above hold for every $\beta > 0$. From now on, unless stated otherwise, $\beta > 1$ is fixed before $n \to \infty$, and all limits are taken as $n \to \infty$. We use the following conventions.

Eventual statements. A statement holds *for all sufficiently large n*, or *eventually*, if there exists n_0 such that it holds for all $n \geq n_0$. The threshold n_0 may depend on β and on the fixed parameters named in the statement, such as an accuracy parameter η or ε or a threshold τ. It never depends on the disorder, on a configuration, or on a time.

Constants. We write $c, C, c_\beta, C_\beta, \dots$ for positive finite constants. They depend only on β and on fixed parameters that are listed explicitly, and never on n, on the disorder, or on configurations. Within a proof, the value of an unnamed constant may change from line to line. Named constants, such as the coefficient A_β of Theorem 1.1, keep their meaning throughout. When several parameters are chosen, we say in which order they are fixed. A statement made "for every fixed τ" asserts nothing uniform in τ.

Power exponents. For a positive sequence (a_n) and $r \in \mathbb{R}$, we write $a_n = n^{r+o(1)}$ if, for every fixed $\eta > 0$, we have $n^{r-\eta} \le a_n \le n^{r+\eta}$ for all sufficiently large n, with a threshold that may depend on η. For example, $\mathrm{Var}(F_n) = n^{1/3+o(1)}$ means that for every fixed $\eta > 0$, eventually $n^{1/3-\eta} \le \mathrm{Var}(F_n) \le n^{1/3+\eta}$. For nonnegative random variables U_n, we write $U_n = n^{r+o_{\mathbb{P}}(1)}$ if $\mathbb{P}\{n^{r-\eta} \le U_n \le n^{r+\eta}\} \to 1$ for every fixed $\eta > 0$. The notation $|F_n - \mathbb{E}F_n| = n^{1/6+o_{\mathbb{P}}(1)}$ means

$$\mathbb{P}\big\{n^{1/6-\eta} \le |F_n - \mathbb{E}F_n| \le n^{1/6+\eta}\big\} \longrightarrow 1 \quad \text{for every fixed } \eta > 0. \tag{3.17}$$

This is a statement about the typical size of $F_n - \mathbb{E}F_n$. It does not follow from $\mathrm{Var}(F_n) = n^{1/3+o(1)}$, since a variance estimate does not exclude that $|F_n - \mathbb{E}F_n|$ is much smaller than $n^{1/6}$ with probability bounded away from zero.

3.6 Normalizations in the cited works

The works from which we import results use several normalizations. Three operations relate them to ours, and they act differently.

(i) Replacing the counting measure on Σ_n by the uniform probability measure replaces Z_n by $2^{-n}Z_n$, F_n by $F_n - n\log 2$, and $p_n(\beta)$ and P_β by $p_n(\beta) - \log 2$ and $P_\beta - \log 2$. This is a deterministic shift. It changes neither Gibbs measures, nor variances, nor the centered variable $F_n - \mathbb{E}F_n$, nor the mean bias $\Delta_{n,\beta}$.

(ii) Completing the Hamiltonian as in Section 3.2 adds the random variable βG to F_n. By Lemma 3.1, it changes neither Gibbs measures nor expected logarithms, but it adds $\beta^2/2$ to the variance, adds $\beta^2 t^2/4$ to the logarithm of the Laplace transform of the centered variable, and multiplies the sth moment of an ℓ-replica partition function by $e^{\beta^2\ell^2 s^2/4}$.

(iii) The energy normalization of [8, Section 7.1] divides log partition functions and the Parisi functional by β and multiplies the cumulative profile by β. It also rescales the spatial variable of the Parisi equation, which changes spatial derivatives of the Parisi solution by powers of β.

Table 3.1 lists the conventions of the works cited most often. In it, Φ_α and $\mathcal{P}_\beta$ are the solution of the Parisi equation and the Parisi functional of Section 4, with terminal value $\Phi_\alpha(1,z) = \log(2\cosh z)$.

OpenAI's upper-bound paper uses the same normalization [85, Section 2]. Its completion matrix $W(G)$ is the matrix W of Section 3.2; its $\widehat{H}_G$, $F(G)$, and π_G are our H_n^c, F_n^c, and π^J, respectively. The diagonal contributes only the common Gaussian variable G of (3.5), so the heat-bath kernels also agree by (3.8). The primitive matrix representation used in Part IV is given in Section 13.1.

We record the consequences for the results of this paper. The barrier preprint [77] uses our normalization, and its results apply without translation. The fluctuation preprint

[81] places β inside the Hamiltonian and uses the uniform prior. Since $\widetilde{F}_n - \mathbb{E}\widetilde{F}_n = F_n - \mathbb{E}F_n$ in the notation of Table 3.1, its fluctuation theorem has the same meaning in our normalization, and its logarithmic-moment bounds concern the variable $F_n^c - \mathbb{E}F_n^c$, whose Laplace transform is related to that of $F_n - \mathbb{E}F_n$ by (3.6). The works [6, 8] use our normalization of the model. The variance–bias inequality of [8, Remark 7.14, (7.68)] and the deduction in [6, Section 1.6] are stated directly for $\operatorname{Var} F_n$ and $\Delta_{n,\beta}$. The additive term $\beta^2/2$ that appears there is the variance of βG from part (c) of Lemma 3.1, not the effect of a change of prior. Neither the completion nor the choice of prior changes any exponent in Theorem 1.6 or Corollary 1.7.

Work	Model and pressure	Completion and other conventions
OpenAI, barriers [77, Section 1, (1.1), pp. 2–3; Section 2, (2.1), (2.12)–(2.14)]	As here: H_n^J, $Z_{n,\beta}^J = \sum_x e^{\beta H_n^J(x)}$, π^J, R, $p_n(\beta)$; the same Φ_α and $\mathcal{P}_\beta$.	Adds $(2n)^{-1/2}\sum_i J_{ii}$, with new independent J_{ii}, in free-energy calculations only; this variable has the law of G. The kernels K^J and P_t^J and the distances d^J, d^J_{disc} are as here.
OpenAI, fluctuations [81, Section 1, p. 1; Section 5.2, p. 48; Section 7, p. 66]	Writing $\widetilde{H}_n$, $\widetilde{Z}_n$, $\widetilde{F}_n$ for the H_n, Z_n, F_n of that work: $g_{ij} = J_{ij}$, $\sigma = x$, $\widetilde{H}_n = \beta H_n$, $\widetilde{Z}_n = 2^{-n}\sum_\sigma e^{\widetilde{H}_n(\sigma)} = 2^{-n} Z_n$, $\widetilde{F}_n = F_n - n\log 2$.	$\widetilde{F}_n^c = \widetilde{F}_n + D$ with $D \sim N(0, \beta^2/2)$, which has the law of βG; hence $\widetilde{F}_n^c - \mathbb{E}\widetilde{F}_n^c$ has the law of $F_n^c - \mathbb{E}F_n^c$. $T = \beta^2$; the terminal value is $\log\cosh z$.
Aronow and Lopatto [8, (1.1)–(1.3), p. 4; Section 7.1, pp. 53–54; (7.64), p. 93]	$N = n$, $g_{ij} = J_{ij}$, $\sigma = x$; $F_{N,\beta} = F_n$ and $p_{N,\beta} = p_n(\beta)$; $a = \alpha$ and $u_a = \Phi_\alpha$, with the same functional $\mathcal{P}_\beta$.	$D_N = \operatorname{Tr} W/2$ has the law of G, and $F_{N,\beta}^c = F_n + \beta D_N$. In Section 7: $\gamma = \beta\alpha$, $u_\gamma^{\mathrm{en}}(q,x) = \beta^{-1}\Phi_\alpha(q,\beta x)$, the functional of γ is $\beta^{-1}\mathcal{P}_\beta(\alpha)$, and $X_{N,\beta} = F_n/\beta$. The lower bound $\gamma \geq \lambda$ on the cumulative profile is $\alpha \geq \lambda/\beta$; the profile $\max\{\alpha, 3s\}$ corresponds to $\lambda = 3\beta s$.
Aronow and Lopatto [6, Section 1.1, p. 1; Section 3.1, p. 33]	As in [8]: $F_{N,\beta} = F_n$, $p_{N,\beta} = p_n(\beta)$, and $u_\alpha = \Phi_\alpha$; α_β is the distribution function of the Parisi measure.	D_N and $F_{N,\beta}^c = F_n + \beta D_N$ as in [8]. The mean bias $P_\beta - p_{N,\beta}$ is $\Delta_{n,\beta}$.

Table 3.1. Normalizations of the works cited most often, compared with (3.1) and (3.5). In the second and third columns, the symbol of the cited work appears on the left of each identity. In [6, 8, 77], P_β is defined as the infimum of $\mathcal{P}_\beta$; it equals our P_β by Theorem 4.4.

SECTION 4

Gaussian hierarchies and imported results

This section fixes the scalar and matrix objects used in Parts II and III, and it states, once, the results from the literature on which the proof of Theorem 1.1 relies. In Section 4.1 we define the Gaussian integration step and record its elementary properties. In Section 4.2 we define the Parisi functional, the scalar diffusion and the response Γ, and we state the Parisi formula, the uniqueness of its minimizer, the scalar estimates and stationarity identities of OpenAI's barrier preprint [77], and the fact that the support of the Parisi measure is an interval. Together these imply that $\Gamma(t) = t$ on $[0, q_\beta]$ (Corollary 4.8). Section 4.3 states the finite-size estimates of [6, 8]. In Section 4.4 we define Gaussian hierarchies, their trial values and the constrained partition functions, we prove the block factorization of the unperturbed three-replica path (Lemma 4.20), and we state the matrix comparison and two Gaussian tools of [77]. Section 4.5 states the two- and three-replica pressure bounds of [77], and Section 4.6 its energy and rotation estimates and its bound on the masses of overlap neighborhoods.

We use the scalar objects from Section 5 on. The hierarchies and the matrix comparison enter Sections 6 to 9, the finite-size estimates enter Sections 8 to 10 and 12, and the bounds on the masses of overlap neighborhoods enter Section 10. We prove Lemmas 4.2 and 4.20 and Corollary 4.8 and the elementary facts recorded in the remarks. Every theorem and proposition of this section is quoted, without proof, from the source named in its heading, or is a consequence of a proof given there; in the latter case we say so and explain how it is obtained (Proposition 4.25). A note after each result lists the sections in which it is used.

Imported results and their normalization. Apart from standard tools of probability, such as Gaussian integration by parts [107, Section A.4, (A.17)], Itô's formula [64, Theorem 3.3.6], Gronwall's inequality [107, Section A.13], and the inequalities of Markov and Chebyshev (apply [107, (A.7)] to a nonnegative variable and to a squared centered variable, respectively), the results stated in this section with a citation are the only mathematical inputs from the literature in the proof of Theorem 1.1. The inputs used only in Part V are stated in Sections 20 and 21. The inputs to Part IV are listed in Section 13.4. The barrier preprint [77] uses our normalization (Table 3.1), and it defines constrained pressures as $n^{-1}\mathbb{E}\log$ of unnormalized constrained sums. Every statement of [77] quoted below holds after the changes of letters listed in Table 4.1, apart from the explicitly

described consequences of proofs in Proposition 4.25. For each of the other sources, we state, where it is quoted, only the change of normalization that the quoted result needs; Section 3.6 compares all of them with ours.

Table 4.1. Notation of [77] and of this paper.

Object	In [77]	Here
number of replicas	d	ℓ
constrained pressure	$F_{d,n}(D)$	$p_{\ell,n}(D)$ (Definition 4.16)
hierarchy masses	$x_{-1}=0$, $x_k=1$, and $x_{j-1}\le x_j$; the jth field is integrated at mass x_{j-1}	$0\le m_1\le\cdots\le m_N\le 1$; the jth increment is integrated at mass m_j (Definition 4.14)
squared norm, penalty	$\lvert U\rvert^2$, $\int x\,\mathrm{d}\lvert Q\rvert^2$	$\lVert U\rVert_F^2$, $\int m\,\mathrm{d}\lVert Q\rVert_F^2$ (4.14)
minimizer	a minimizer μ	μ_β, unique by Theorem 4.5
scalar diffusion	X_s	Y_t
scalar quantities	$A(s)$, $d_*(s)$	$\int_s^1 \alpha(t)\,\mathrm{d}t$, $s\alpha(s)+\int_s^1\alpha(t)\,\mathrm{d}t$
large-overlap cutoff in [77, Lemma 4.1 and (4.20)]	h	b_*
scales in [77, Sections 6–7]	$L=n^{1/10000}$, $\rho=L^{-1}$	L_{OB}, ρ_{OB}

Conventions. Throughout, we fix $\beta>0$. Every statement that involves the Parisi measure μ_β, its support $[0,q_\beta]$, or the response Γ of the minimizer assumes $\beta>1$. Constants C,c and thresholds n_0 depend only on β and on the parameters named in the statement where they appear.

4.1 Gaussian integration steps

In this section, $\ell\ge 1$ is an integer. We write vectors of $\mathbb{R}^\ell$ in bold, $\mathbf{z}\cdot\mathbf{y}$ and $\lvert\mathbf{z}\rvert$ for the Euclidean inner product and norm, and $\mathbf{g}$ for a standard Gaussian vector in $\mathbb{R}^\ell$. The notation $\Sigma\succeq 0$ means that the symmetric $\ell\times\ell$ matrix Σ is positive semidefinite, and $\Sigma^{1/2}$ denotes its positive semidefinite square root. A function $f\colon\mathbb{R}^\ell\to\mathbb{R}$ has *linear growth* if it is measurable and there exists $C<\infty$ such that $\lvert f(\mathbf{z})\rvert\le C(1+\lvert\mathbf{z}\rvert)$ for all $\mathbf{z}$.

Definition 4.1 (Gaussian integration step)**.** Let $m\in[0,1]$, let $\Sigma\succeq 0$ be an $\ell\times\ell$ matrix, and let f have linear growth. The *Gaussian integration step* at mass m with covariance Σ is

$$T_{m,\Sigma}f(\mathbf{z})=\begin{cases} m^{-1}\log\mathbb{E}\exp\bigl(mf(\mathbf{z}+\Sigma^{1/2}\mathbf{g})\bigr), & m>0,\\ \mathbb{E}f(\mathbf{z}+\Sigma^{1/2}\mathbf{g}), & m=0. \end{cases}\tag{4.1}$$

The *tilted Gaussian kernel* is the probability kernel

$$\mathcal{K}_{m,\Sigma,f}(\mathbf{z},\, \mathrm{d}\mathbf{y}) = \exp\bigl\{m\bigl[f(\mathbf{y}) - T_{m,\Sigma}f(\mathbf{z})\bigr]\bigr\}\, N(\mathbf{z},\Sigma)(\,\mathrm{d}\mathbf{y}), \tag{4.2}$$

where $N(\mathbf{z},\Sigma)$ is the law of $\mathbf{z}+\Sigma^{1/2}\mathbf{g}$. For $m=0$ the exponential factor equals one.

Note that $T_{m,\Sigma}f$ is the normalization that makes $\mathcal{K}_{m,\Sigma,f}(\mathbf{z},\cdot)$ a probability measure. A singular Σ is allowed. In that case, $N(\mathbf{z},\Sigma)$ is carried by $\mathbf{z}+\operatorname{range}\Sigma$, and no inverse of Σ is used. A rank-one covariance $\Sigma=\mathbf{v}\mathbf{v}^{\mathsf{T}}$ integrates a single Gaussian variable in the direction $\mathbf{v}$, and $T_{m,0}f=f$.

Lemma 4.2 (Elementary properties of the integration step)**.** *Let $m\in[0,1]$, let $\Sigma,\Sigma'\succeq 0$ be $\ell\times\ell$ matrices, and let f, f' have linear growth.*

(a) *The function $T_{m,\Sigma}f$ has linear growth. If $f\le f'$, then $T_{m,\Sigma}f\le T_{m,\Sigma}f'$, and $T_{m,\Sigma}(f+c)=T_{m,\Sigma}f+c$ for every $c\in\mathbb{R}$.*

(b) *$T_{m,\Sigma}T_{m,\Sigma'}f=T_{m,\Sigma+\Sigma'}f$.*

(c) *Let $\ell=\ell_1+\ell_2$, write $\mathbf{z}=(\mathbf{z}_1,\mathbf{z}_2)$ accordingly, and suppose that $f(\mathbf{z})=f_1(\mathbf{z}_1)+f_2(\mathbf{z}_2)$, with f_1, f_2 of linear growth, and that Σ is block diagonal with blocks Σ_1,Σ_2. Then $T_{m,\Sigma}f(\mathbf{z})=T_{m,\Sigma_1}f_1(\mathbf{z}_1)+T_{m,\Sigma_2}f_2(\mathbf{z}_2)$.*

(d) *For every integer $k\ge 1$, we have $T_{m/k,\Sigma}(kf)=k\,T_{m,\Sigma}f$.*

(e) *The value $T_{m,\Sigma}f(\mathbf{z})$ depends only on the restriction of f to $\mathbf{z}+\operatorname{range}\Sigma$.*

Proof. For (a), suppose that $|f(\mathbf{z})|\le C(1+|\mathbf{z}|)$, and set $M=C\|\Sigma^{1/2}\||\mathbf{g}|$, where $\|\cdot\|$ is the operator norm. Then $|f(\mathbf{z}+\Sigma^{1/2}\mathbf{g})|\le C(1+|\mathbf{z}|)+M$. For $m>0$, by Jensen's inequality we have $T_{m,\Sigma}f(\mathbf{z})\ge \mathbb{E}f(\mathbf{z}+\Sigma^{1/2}\mathbf{g})\ge -C(1+|\mathbf{z}|)-\mathbb{E}M$. By Hölder's inequality, the map $m\mapsto m^{-1}\log\mathbb{E}e^{mM}$ is nondecreasing on $(0,1]$, which implies that $T_{m,\Sigma}f(\mathbf{z})\le C(1+|\mathbf{z}|)+\log\mathbb{E}e^{M}$. The case $m=0$ is the same with expectations, and the monotonicity and the translation identity are immediate from (4.1).

For (b) with $m>0$, we have $\exp(mT_{m,\Sigma'}f(\mathbf{y}))=\mathbb{E}\exp(mf(\mathbf{y}+\Sigma'^{1/2}\mathbf{g}'))$ for an independent standard Gaussian vector $\mathbf{g}'$. Applying $T_{m,\Sigma}$, we find

$$\exp\bigl(mT_{m,\Sigma}T_{m,\Sigma'}f(\mathbf{z})\bigr)=\mathbb{E}\exp\bigl(mf(\mathbf{z}+\Sigma^{1/2}\mathbf{g}+\Sigma'^{1/2}\mathbf{g}')\bigr),$$

and $\Sigma^{1/2}\mathbf{g}+\Sigma'^{1/2}\mathbf{g}'$ is centered Gaussian with covariance $\Sigma+\Sigma'$. The case $m=0$ is the same computation without exponentials. For (c), if Σ is block diagonal, then the two blocks of $\Sigma^{1/2}\mathbf{g}$ are independent, and the expectation of $e^{mf_1}e^{mf_2}$ factorizes (for $m=0$, the expectation of f_1+f_2 splits). Part (d) follows from (4.1), since $(k/m)\log\mathbb{E}e^{(m/k)kf}=k\,m^{-1}\log\mathbb{E}e^{mf}$ for $m>0$. Part (e) holds because $\mathbf{z}+\Sigma^{1/2}\mathbf{g}\in\mathbf{z}+\operatorname{range}\Sigma$ almost surely. □

Used in: Lemma 4.20 below, and Sections 6 to 8, in the computations with Gaussian hierarchies.

4.2 The Parisi functional and the scalar diffusion

A *distribution function* is a right-continuous nondecreasing function $\alpha\colon [0,1] \to [0,1]$ such that $\alpha(1) = 1$. Equivalently, $\alpha(t) = \mu([0,t])$ for a probability measure μ on $[0,1]$. We pass freely between α and μ. We say that α is a *step* distribution function if there exist $0 = t_0 < t_1 < \cdots < t_p = 1$ such that α is constant on each interval $[t_{i-1}, t_i)$. We call such points *breakpoints* of α, and $m_i = \alpha(t_{i-1})$ denotes the value of α on $[t_{i-1}, t_i)$.

For a step distribution function α, we define $\Phi_\alpha\colon [0,1] \times \mathbb{R} \to \mathbb{R}$ by $\Phi_\alpha(1,z) = \log(2\cosh z)$ and, for $i = p, p-1, \dots, 1$ and $t_{i-1} \le t < t_i$,

$$\Phi_\alpha(t,\cdot) = T_{m_i,\beta^2(t_i-t)}\Phi_\alpha(t_i,\cdot), \tag{4.3}$$

with the scalar integration step of Definition 4.1 ($\ell = 1$). On each interval (t_{i-1}, t_i), the function $e^{m_i\Phi_\alpha}$ (or Φ_α itself, if $m_i = 0$) solves the backward heat equation $-\partial_t\psi = (\beta^2/2)\partial_{zz}\psi$. It follows that Φ_α is smooth there and solves the Parisi equation

$$-\partial_t\Phi_\alpha(t,z) = \frac{\beta^2}{2}\Big(\partial_{zz}\Phi_\alpha(t,z) + \alpha(t)\big(\partial_z\Phi_\alpha(t,z)\big)^2\Big), \qquad \Phi_\alpha(1,z) = \log(2\cosh z). \tag{4.4}$$

For a general distribution function α, we define Φ_α as the uniform limit of Φ_{α_k} along any sequence of step distribution functions α_k such that $\int_0^1 |\alpha_k - \alpha|\,\mathrm{d}t \to 0$. By Proposition 4.3 below, the limit exists and does not depend on the sequence, and (4.4) holds in time-integrated form, with the time derivative existing for almost every t. The *Parisi functional* is

$$\mathcal{P}_\beta(\alpha) = \Phi_\alpha(0,0) - \frac{\beta^2}{2}\int_0^1 t\alpha(t)\,\mathrm{d}t. \tag{4.5}$$

We write $u_\alpha = \partial_z\Phi_\alpha$ and $\chi_\alpha = \partial_{zz}\Phi_\alpha$, and we define the scalar diffusion and the response by

$$\mathrm{d}Y_t = \beta\,\mathrm{d}B_t + \beta^2\alpha(t)u_\alpha(t,Y_t)\,\mathrm{d}t, \qquad Y_0 = 0, \qquad \Gamma_\alpha(t) = \mathbb{E}u_\alpha(t,Y_t)^2, \tag{4.6}$$

where B is a standard Brownian motion. These are the scalar conventions of [77, Section 3], where the diffusion is denoted X.

Proposition 4.3 (Regularity, stability, and scalar evolution [77, Lemma 3.1])**.** *Let $\beta > 0$. There exists $C_\beta < \infty$ such that the following holds for all distribution functions α and $\widetilde{\alpha}$.*

(i) *The limit defining Φ_α exists and does not depend on the approximating sequence. The functions Φ_α, u_α and χ_α are continuous on $[0,1]\times\mathbb{R}$, $\Phi_\alpha(t,\cdot)$ is even, and*

$$|u_\alpha| \le 1, \qquad 0 < \chi_\alpha \le 1, \qquad |\chi_\alpha(t,z) - \chi_\alpha(t,z')| \le C_\beta|z - z'|.$$

(ii) *We have*

$$\sup_{t,z}\Big(|\Phi_\alpha - \Phi_{\widetilde{\alpha}}| + |u_\alpha - u_{\widetilde{\alpha}}| + |\chi_\alpha - \chi_{\widetilde{\alpha}}|\Big) \le C_\beta\int_0^1 |\alpha(t) - \widetilde{\alpha}(t)|\,\mathrm{d}t.$$

(iii) *The equation* (4.6) *has a unique strong solution. If α is a step distribution function with breakpoints $t_0 < \cdots < t_p$, then $(Y_{t_i})_{0\le i\le p}$ is a Markov chain whose transition from t_{i-1} to t_i is the tilted Gaussian kernel $\mathcal{K}_{m_i,\beta^2(t_i-t_{i-1}),\Phi_\alpha(t_i,\cdot)}$.*

(iv) *For $0 \le s \le t \le 1$,*

$$u_\alpha(t, Y_t) - u_\alpha(s, Y_s) = \beta \int_s^t \chi_\alpha(r, Y_r)\, \mathrm{d}B_r,$$

$$\mathbb{E}\big[\chi_\alpha(t, Y_t) \mid Y_s\big] = \chi_\alpha(s, Y_s) - \beta^2 \int_s^t \alpha(r)\mathbb{E}\big[\chi_\alpha(r, Y_r)^2 \mid Y_s\big]\, \mathrm{d}r.$$

Moreover, Γ_α is continuously differentiable on $[0,1]$, with one-sided derivatives at the endpoints, and $\Gamma_\alpha(0) = 0$, $\Gamma_\alpha'(t) = \beta^2\mathbb{E}\chi_\alpha(t, Y_t)^2$.

(v) *Let $0 \le a \le 1$ and $z \in \mathbb{R}$, and let $Y^{a,z}$ solve $\mathrm{d}Y_t^{a,z} = \beta\, \mathrm{d}B_t + \beta^2\alpha(t)u_\alpha(t, Y_t^{a,z})\, \mathrm{d}t$ for $a \le t \le 1$, with $Y_a^{a,z} = z$. This equation has a unique strong solution, and the two identities in* (iv) *hold for $Y^{a,z}$ and $a \le s \le t \le 1$. If α is a step distribution function, then the transition of $Y^{a,z}$ across each interval of constancy of α after a, with a added to the breakpoints, is the tilted Gaussian kernel of* (iii)*. The function $\Gamma_\alpha^{a,z}(t) = \mathbb{E}u_\alpha(t, Y_t^{a,z})^2$ is continuously differentiable on $[a,1]$, with $\Gamma_\alpha^{a,z}(a) = u_\alpha(a,z)^2$ and $(\Gamma_\alpha^{a,z})'(t) = \beta^2\mathbb{E}\chi_\alpha(t, Y_t^{a,z})^2$.*

We note that, in [77], part (i) collects (3.2) and the evenness of Φ_α (so $u_\alpha(t,0) = 0$), part (ii) is (3.3), part (iii) is the statement about (3.4), and part (iv) contains (3.5), (3.7), and (3.8). Taking expectations in the second identity of (iv), we obtain the unconditional form [77, (3.6)]. The last sentence of [77, Lemma 3.1] extends the well-posedness and the identities (3.5)–(3.6) to a diffusion started at any deterministic time and position. In part (v), the conditional identity then follows from the Markov property. The statement about the transition kernels follows from the identification of the kernels on each interval of constant mass in the proof of [77, Lemma 3.1], which starts from an arbitrary position. The statement about $\Gamma_\alpha^{a,z}$ follows from the martingale identity, Itô's isometry, and the continuity of χ_α and $Y^{a,z}$. The initial value $\Gamma_\alpha(0) = 0$ in (iv) uses the start $Y_0 = 0$. The constant C_β does not depend on α. The proof in [77] also bounds $|\partial_z^3\Phi_\alpha|$ by $2e^{3\beta^2}$ for step distribution functions, and it proves continuity in β. We prove bounds on all positive spatial derivatives, for every distribution function, in Section 5. For the step approximation and the spatial regularity of the Parisi evolution, [77] credits Auffinger and Chen [12, Propositions 1–2].

Used in: Section 5, for all scalar estimates, and Sections 6 to 9, for the stability under a change of the distribution function and for the passage from step approximations to the minimizer, and Part IV, for the scalar process and its stochastic identities.

Theorem 4.4 (Guerra's bound and the Parisi formula [59, Theorem 3], [60, Theorems 1–2], [103, Theorem 1.1])**.** *Let $\beta > 0$. For every $n \ge 2$ and every distribution function α, we have $p_n(\beta) \le \mathcal{P}_\beta(\alpha)$. The limit $P_\beta = \lim_{n\to\infty} p_n(\beta)$ exists, and $P_\beta = \inf_\alpha \mathcal{P}_\beta(\alpha)$, where the infimum is over all distribution functions. In particular, $\Delta_{n,\beta} = P_\beta - p_n(\beta) \ge 0$ for every $n \ge 2$.*

The bound $p_n(\beta) \le \mathcal{P}_\beta(\alpha)$ is Guerra's replica-symmetry-breaking bound [59, Theorem 3], whose proof first treats step distribution functions and extends to the full class by [59, Theorem 1]. In our normalization it is [77, (2.14)], where it is deduced from Proposition 4.19 below with $\ell = 1$ (see Lemma 4.20(a)) and extended to all α by Proposition 4.3(ii). The existence of the limit was proved by Guerra and Toninelli [60, Theorems 1–2] by an interpolation that shows superadditivity. The identification of the limit is Talagrand's Parisi formula [103, Theorem 1.1, (1.14)], applied at zero external field with counting measure and covariance function $\xi(q) = \beta^2 q^2/2$.

Used in: Section 3, for the identification of P_β with the infimum of $\mathcal{P}_\beta$, and Sections 7 to 9, where P_β is the baseline of every pressure comparison, and Part IV, for the finite-volume pressure bound.

Theorem 4.5 (Uniqueness of the Parisi measure [13, Theorem 2 and Corollary 1])**.** *Let $\beta > 0$. The functional $\mathcal{P}_\beta$ has exactly one minimizer among distribution functions.*

A minimizer exists by an elementary compactness argument. The probability measures on $[0,1]$ form a weakly compact set, weak convergence implies L^1 convergence of the distribution functions, and $\mathcal{P}_\beta$ is continuous in L^1 by Proposition 4.3(ii), as in [77, proof of Proposition 3.2]. The functional of Auffinger and Chen [13, (7)] is $\mathcal{P}^{\mathrm{AC}}_{\xi,h}(\mu) = \Phi^{\mathrm{AC}}_\mu(0,h)$, where Φ^{AC}_μ solves the Parisi equation with coefficient ξ'' and terminal value $\log\cosh$. For $\xi(q) = \beta^2 q^2/2$ and $h = 0$, we have $\xi'' = \beta^2$, and Remark 4.9 gives

$$\mathcal{P}_\beta(\alpha) = \log 2 + \mathcal{P}^{\mathrm{AC}}_{\xi,0}(\mu) - \frac{\beta^2}{2}\int_0^1 t\alpha(t)\,\mathrm{d}t,$$

which is their full variational objective [13, (8)]. They prove that $\mathcal{P}^{\mathrm{AC}}_{\xi,0}$ is strictly convex [13, Theorem 2]. Subtracting the linear functional displayed above preserves strict convexity, and [13, Corollary 1] gives uniqueness of the minimizer. We denote the minimizer by α_β and the corresponding probability measure, the *Parisi measure*, by μ_β. By uniqueness, we may speak of the scalar data of the minimizer; the results of [77] quoted below are stated there for an arbitrary minimizer.

Used in: every section of Parts II to IV.

Standing notation. Let $\beta > 1$. From now on we write $\alpha = \alpha_\beta$ and $\mu = \mu_\beta$ when no confusion can arise, and Φ, u, χ, Y, and Γ denote Φ_α, u_α, χ_α, the diffusion (4.6) of α, and Γ_α. We also write $w = \partial_z^3\Phi$; this derivative exists and is bounded by the estimates of Section 5. Some sections of Part II abbreviate $b = \beta^2$. In this notation, (4.4) and (4.6) read

$$\begin{aligned} \partial_t\Phi &= -\frac{b}{2}\big(\partial_{zz}\Phi + \alpha(\partial_z\Phi)^2\big), \qquad \Phi(1,z) = \log(2\cosh z), \\ \mathrm{d}Y_t &= b\alpha(t)u(t,Y_t)\,\mathrm{d}t + \sqrt{b}\,\mathrm{d}B_t, \qquad Y_0 = 0. \end{aligned} \tag{4.7}$$

Proposition 4.6 (Support conditions and the mean of χ [77, Proposition 3.2])**.** *Let $\beta > 0$, and let α be a minimizer of $\mathcal{P}_\beta$, with probability measure μ, diffusion Y and response Γ.*

(i) *We have $0 \in \operatorname{supp}\mu$, and $\operatorname{supp}\mu \subset [0, 1-\varepsilon_0]$ for some $\varepsilon_0 > 0$.*

(ii) *At every $q \in \operatorname{supp}\mu \cap (0,1)$, we have $\Gamma(q) = q$ and $\Gamma'(q) \le 1$. Also $\Gamma(0) = 0$ and $\Gamma'(0) \le 1$.*

(iii) *For every $s \in [0,1]$,*

$$\mathbb{E}\big[\chi(s, Y_s) + \alpha(s)u(s, Y_s)^2\big] = s\alpha(s) + \int_s^1 \alpha(t)\,\mathrm{d}t, \tag{4.8}$$

or equivalently

$$\mathbb{E}\chi(s, Y_s) = \int_s^1 \alpha(t)\,\mathrm{d}t + \alpha(s)\big(s - \Gamma(s)\big). \tag{4.9}$$

In particular, $\chi(0,0) = \int_0^1 \alpha(t)\,\mathrm{d}t$ and $\beta\int_0^1 \alpha(t)\,\mathrm{d}t \le 1$. Moreover, $\alpha(s) > 0$ for every $s \in (0,1]$.

In [77], the right side of (4.8) is written $d_*(s) = 1 - \int_{(s,1]} t\,\mu(\mathrm{d}t)$, which equals $s\alpha(s) + \int_s^1 \alpha(t)\,\mathrm{d}t$ by Fubini's theorem, and $\int_s^1 \alpha(t)\,\mathrm{d}t$ is written $A(s)$. The identities (4.8) and (4.9) are [77, (3.14)–(3.15)], where they are called stationarity identities. They hold at every s, including atoms of μ. In this paper, the name *stationarity identities* refers instead to the two identities (5.15) on $(0, q_\beta)$ described after Corollary 4.8. The support conditions in (ii) come from the first variation [77, (3.9)]. As credited in [77, Section 3], the first-variation characterization appears in Chen [36, Theorem 2 and Proposition 1] and in Jagannath and Tobasco [63, Proposition 1.1 and Corollary 3.10]. Part (ii) does not describe the support of μ; for $\beta > 1$, Theorem 4.7 describes it.

Used in: Corollary 4.8, Section 5, and Section 9, where (4.9) bounds $\mathbb{E}\chi(q, Y_q) - \int_q^1 \alpha(t)\,\mathrm{d}t$ at overlaps q with $|q - \Gamma(q)|$ small.

Theorem 4.7 (Full replica symmetry breaking [69, Theorem 1.1])**.** *For every $\beta > 1$, there exists $q_\beta \in (0,1)$ with the following properties. We have $\operatorname{supp}\mu_\beta = [0, q_\beta]$. The restriction of μ_β to $[0, q_\beta)$ has a nonnegative density with respect to Lebesgue measure that is infinitely differentiable on $(0, q_\beta)$ and whose derivatives of every order extend continuously to 0. Finally, $\mu_\beta(\{q_\beta\}) \in (0,1)$.*

[69] works with the Hamiltonian $\beta(2n)^{-1/2}\sum_{i,j=1}^n g_{ij}x_ix_j$, with (g_{ij}) independent standard Gaussians, which equals βH_n plus an independent centered Gaussian of variance $\beta^2/2$ common to all configurations, that is, βH_n^c with H_n^c the completed Hamiltonian of Section 3.2. The Parisi functional there is $\log 2 + u_\mu(0,0) - (\beta^2/2)\int_0^1 s\mu([0,s])\,\mathrm{d}s$, where u_μ solves the Parisi equation with terminal value $\log\cosh x$. By Remark 4.9, this functional equals $\mathcal{P}_\beta$, and μ_β is the same measure in both papers. We cite version 3 of [69].

Its version 4 (October 2026) states the same support conclusion, with a density that is smooth up to q_β, and adds a quantitative bound on the atom. We use only the statement above.

Used in: Corollary 4.8, and through it the stationarity identities of Section 5 and the choice of the intervals of overlaps in Sections 7, 9, and 10, and Part IV, for the continuation of the scalar process beyond the support.

Corollary 4.8 (The interval on which $\Gamma(t) = t$)**.** *Let $\beta > 1$. Then*

$$\Gamma(t) = t \qquad (0 \le t \le q_\beta), \tag{4.10}$$

and $\beta^2 \mathbb{E}\chi(t, Y_t)^2 = 1$ for $0 < t < q_\beta$. Moreover, $\alpha_\beta(0) = 0$, $\alpha_\beta(t) > 0$ for $0 < t \le 1$, $\alpha_\beta(t) = 1$ for $t \ge q_\beta$, and α_β is infinitely differentiable on $[0, q_\beta)$, with one-sided derivatives at 0.

Proof. By Theorem 4.7, we have $\operatorname{supp} \mu_\beta = [0, q_\beta]$ with $q_\beta < 1$. Then $\operatorname{supp} \mu_\beta \cap (0, 1) = (0, q_\beta]$, and Proposition 4.6(ii) implies that $\Gamma(q) = q$ for $0 < q \le q_\beta$. Also $\Gamma(0) = 0$ by Proposition 4.3(iv), which proves (4.10). Recall that Γ is continuously differentiable with $\Gamma'(t) = \beta^2 \mathbb{E}\chi(t, Y_t)^2$. Differentiating (4.10) on $(0, q_\beta)$, we obtain the second claim. For $t < q_\beta$, Theorem 4.7 implies that $\alpha_\beta(t) = \mu_\beta([0, t])$ is the integral from 0 to t of a density that is smooth on $[0, q_\beta)$; in particular $\alpha_\beta(0) = 0$. For $t \ge q_\beta$ we have $\alpha_\beta(t) = 1$. Finally, $\alpha_\beta(t) > 0$ for $t > 0$ because $0 \in \operatorname{supp} \mu_\beta$. □

In Section 5, we prove that the identity $\beta^2 \mathbb{E}\chi(t, Y_t)^2 = 1$ and its companion $\mathbb{E}w(t, Y_t)^2 = 2\alpha(t)\mathbb{E}\chi(t, Y_t)^3$ hold on $(0, q_\beta)$ with the regularity needed later. These are the stationarity identities (5.15).

Used in: Section 5, for the stationarity identities, and Sections 7 and 9, for the local comparisons, which take place in $(0, q_\beta)$, and Part IV, for the pre-support scalar identities.

Remark 4.9 (Terminal value $\log\cosh$). Some of the sources quoted here and in Part V, among them OpenAI's fluctuation preprint [81] and [69], use the terminal value $\log\cosh z$ in (4.4). Since $\log(2\cosh z) = \log 2 + \log\cosh z$, the two solutions differ by the constant $\log 2$ (by Lemma 4.2(a) for step distribution functions, and by approximation in general). They have the same spatial derivatives, the same diffusion (4.6) and the same response Γ. Adding $\log 2$ to the functional with terminal value $\log\cosh$ recovers (4.5), and the minimizer does not change. The papers [6, 8] use exactly (4.4) and (4.5).

4.3 Finite-size estimates

We use three estimates of Aronow and Lopatto on the convergence in the Parisi formula at a fixed temperature $\beta > 1$: an upper bound on the variance of F_n, an upper bound on the mean bias $\Delta_{n,\beta} = P_\beta - p_n(\beta)$, and a bound on the increase of the Parisi functional when a lower bound is imposed on the distribution function.

Theorem 4.10 (Upper bound on the variance [8, Theorem 1.4, (1.11)])**.** *For every fixed $\beta > 1$, there exist $C_v < \infty$ and $n_0(\beta)$ such that, for every $n \geq n_0(\beta)$,*

$$\operatorname{Var}(F_n) \leq C_v n^{7/15}. \tag{4.11}$$

The normalization of [8, (1.1)–(1.2)] is ours (Table 3.1). The theorem in [8] also contains the lower bound $\operatorname{Var}(F_n) \geq c_\beta n^{4/15}$ and the bounds $c_\beta n^{-11/12} \leq P_\beta - p_n(\beta) \leq C_\beta n^{-2/3}$ on the mean bias, which we do not use. The upper bound on the mean bias that we use is Theorem 4.11.

Used in: Section 9, for the lower bound on Z_n in the disorder event of Lemma 9.7; and Section 12.7, for the same lower bound in Lemma 12.7.

Theorem 4.11 (Upper bound on the mean bias with exponent 7/10 [6, Theorem 1.1])**.** *For every fixed $\beta > 1$, there exists $C_b < \infty$ such that*

$$0 \leq P_\beta - p_n(\beta) \leq C_b n^{-7/10} \qquad (n \geq 2).$$

The normalization of [6, Section 1.1] is ours, and the lower bound $0 \leq P_\beta - p_n(\beta)$ is Theorem 4.4.

Used in: Section 9, for the lower bound on Z_n in the disorder event of Lemma 9.7, together with Theorem 4.10; Section 10, where it is the input (10.11) for Theorem 1.1, and its constant enters the choice of the scale coefficient in (10.15); and Section 12, through these choices.

Proposition 4.12 (The profile $\max\{\alpha, 3s\}$ [8, Lemma 7.2, (7.3), and Proposition 7.12])**.** *Fix $\beta > 1$ and let $\alpha = \alpha_\beta$. There exist constants $T_* > 0$, $0 < c_1 \leq c_2 < \infty$, $s_0 \in (0, 1/3]$, and $C < \infty$, depending only on β, such that the following holds.*

(i) *For $0 \leq t \leq T_*$, we have $c_1 t \leq \alpha(t) \leq c_2 t$.*

(ii) *For $0 < s \leq s_0$, the function $\max\{\alpha, 3s\}$ is a distribution function, and*

$$0 \leq \mathcal{P}_\beta(\max\{\alpha, 3s\}) - P_\beta \leq C s^5. \tag{4.12}$$

Remark 4.13 (Translation and consequences of Proposition 4.12). In [8, Section 7.1], the scalar problem is rescaled by $\gamma = \beta\alpha$ and $x_{\mathrm{old}} = \beta x$, with rescaled functional $\mathcal{P}^{\mathrm{en}}_\beta(\gamma) = \beta^{-1}\mathcal{P}_\beta(\gamma/\beta)$. By Lemma 7.2 there, applied with temperature threshold $\beta_0 = \beta$ to the rescaled minimizer $\gamma_\beta = \beta\alpha_\beta$, we have $kq \leq \gamma_\beta(q) \leq C_0 q$ on $[0, T_*]$, which is (i) with $c_1 = k/\beta$ and $c_2 = C_0/\beta$. For a lower bound $a > 0$ on the rescaled profile (denoted λ there), its estimate (7.3) bounds the value of the particular trial profile $\max\{a, \gamma_\beta\}$ by $\mathcal{P}^{\mathrm{en}}_\beta(\gamma_\beta) + C_1 a^5$ for $0 < a \leq a_*$, where $a_* > 0$ depends only on β (denoted λ_* there). With $a = 3\beta s$, we have $\max\{a, \gamma_\beta\} = \beta \max\{3s, \alpha\}$, and we find that

$$\mathcal{P}_\beta(\max\{\alpha, 3s\}) = \beta\, \mathcal{P}^{\mathrm{en}}_\beta(\max\{3\beta s, \gamma_\beta\}) \leq P_\beta + 3^5 C_1 \beta^6 s^5 \qquad (3\beta s \leq a_*),$$

which is (ii). The same estimate, in the original variables, is [8, (7.65)–(7.66)] with imposed lower bound $3s$. The bound (4.12) concerns this particular profile, not only the minimum over profiles with an atom at zero. For the minimizer of that constrained problem, Pei [95, Theorem 1(a) and Proposition 2.3(c)] obtains sharp fifth-order asymptotics, which we do not use.

Part (i) localizes the modification. If in addition $3s \le c_1 T_*$, then $\max\{\alpha, 3s\} = \alpha$ on $[3s/c_1, 1]$, since $\alpha(t) \ge c_1 t \ge 3s$ for $3s/c_1 \le t \le T_*$ and $\alpha(t) \ge \alpha(T_*) \ge c_1 T_* \ge 3s$ for $t > T_*$. On $[0, 3s/c_1)$ we have $0 \le \max\{\alpha, 3s\} - \alpha \le 3s$. We conclude that $\int_0^1 |\max\{\alpha, 3s\} - \alpha| \, \mathrm{d}t \le 9s^2/c_1$. The lower bound $3s$ ensures that the mass $\max\{\alpha, 3s\}/3$ of a common block of three coordinates is at least s (Lemma 4.20(b)), as required in Section 8.

Used in: Section 8, for the profile used in the positive fractional comparison.

4.4 Gaussian hierarchies and constrained pressures

We now define the finite Gaussian hierarchies that serve as trials in the pressure comparisons of Part II. Their scalar case is the recursion (4.3). For $\ell \times \ell$ matrices U, V we write $U : V = \sum_{a,b} U_{ab} V_{ab}$ and $\|U\|_F^2 = U : U$.

Definition 4.14 (Gaussian hierarchies)**.** Fix $\ell \ge 1$.

(a) An *admissible path* of length $N \ge 1$ is a pair (Q, m), where $Q = (Q_0, \dots, Q_N)$ are symmetric $\ell \times \ell$ matrices with $Q_0 = 0$ and $\Delta Q_j = Q_j - Q_{j-1} \succeq 0$ for $1 \le j \le N$, and $m = (m_1, \dots, m_N)$ satisfies $0 \le m_1 \le m_2 \le \cdots \le m_N \le 1$. We call Q_N the *endpoint* of the path and m_j the *mass* of the jth increment.

(b) For a terminal function f of linear growth, the *partial values* are $\Psi_N = f$ and

$$\Psi_{j-1} = T_{m_j, \beta^2 \Delta Q_j} \Psi_j \qquad (j = N, N-1, \dots, 1), \tag{4.13}$$

and we set $\Psi(Q, m; f) = \Psi_0$. For a symmetric $\ell \times \ell$ matrix λ, we define

$$f_\lambda(\mathbf{z}) = \log \sum_{\epsilon \in \{-1,1\}^\ell} \exp\bigl(\mathbf{z} \cdot \epsilon + \epsilon^{\mathsf{T}} \lambda \epsilon\bigr), \qquad \Psi(Q, m, \lambda) = \Psi(Q, m; f_\lambda)(0).$$

In particular, $f_0(\mathbf{z}) = \sum_{a=1}^{\ell} \log(2 \cosh z_a)$.

(c) The *path law* of (Q, m, λ) is the law of $(\mathbf{z}_1, \dots, \mathbf{z}_N, \epsilon)$ defined as follows. Let $\mathbf{z}_0 = 0$. Given $\mathbf{z}_0, \dots, \mathbf{z}_{j-1}$, the vector $\mathbf{z}_j$ has law $\mathcal{K}_{m_j, \beta^2 \Delta Q_j, \Psi_j}(\mathbf{z}_{j-1}, \cdot)$, with the partial values of $f = f_\lambda$. Given $\mathbf{z}_1, \dots, \mathbf{z}_N$, the vector $\epsilon \in \{-1, 1\}^\ell$ equals ϵ' with probability $\exp(\mathbf{z}_N \cdot \epsilon' + \epsilon'^{\mathsf{T}} \lambda \epsilon' - f_\lambda(\mathbf{z}_N))$. For $0 \le l \le N$, the *pair law* sharing the initial segment through level l is the law of two such sequences that coincide in $(\mathbf{z}_1, \dots, \mathbf{z}_l)$ and whose continuations, including the terminal draws, are conditionally independent given

$(\mathbf{z}_1, \dots, \mathbf{z}_l)$, each with the transitions above. We write $\mathbb{E}_l^{(2)}$ for the corresponding expectation.

Note that the jth increment is integrated at mass m_j, that mass zero means an ordinary Gaussian expectation, and that the terminal Gibbs draw of ϵ plays the role of a level of mass one. Equivalently, one draws independent increments $\mathbf{z}_j - \mathbf{z}_{j-1}$ with law $N(0, \beta^2 \Delta Q_j)$ and tilts the jth by the density $\exp\{m_j[\Psi_j(\mathbf{z}_j) - \Psi_{j-1}(\mathbf{z}_{j-1})]\}$ given the past. This is the law of [77, (2.5)–(2.6)], in which the continuations of a common initial segment are conditionally independent. By Lemma 4.2(a), every partial value has linear growth, and all quantities above are finite. The *penalty* of an admissible path is

$$\int m \, \mathrm{d}\|Q\|_F^2 = \sum_{j=1}^{N} m_j \big(\|Q_j\|_F^2 - \|Q_{j-1}\|_F^2 \big). \tag{4.14}$$

By Lemma 4.2(b), merging two consecutive increments with the same mass, or inserting a zero increment whose mass keeps the masses ordered, changes neither $\Psi(Q, m, \lambda)$ nor the penalty. Section 8 uses a variant of Definition 4.14 in which all masses are at least a positive number and the disorder is also integrated at positive mass.

Definition 4.15 (Trial values)**.** Let D and λ be symmetric $\ell \times \ell$ matrices, and let (Q, m) be an admissible path. The *trial value* of (Q, m, λ) for D is

$$\mathcal{T}_D(Q, m, \lambda) = \Psi(Q, m, \lambda) - \lambda : D + \frac{\beta^2}{4} \Big(\|D - Q_N\|_F^2 - \int m \, \mathrm{d}\|Q\|_F^2 \Big). \tag{4.15}$$

If $(Q^{(k)}, m^{(k)})$ is a sequence of admissible paths along which the endpoints, the values $\Psi(Q^{(k)}, m^{(k)}, \lambda)$, and the penalties converge, we also call the limit of $\mathcal{T}_D(Q^{(k)}, m^{(k)}, \lambda)$ a trial value for D.

When the endpoint equals D and $\lambda = 0$, the trial value is $\Psi(Q, m, 0) - (\beta^2/4) \int m \, \mathrm{d}\|Q\|_F^2$. This is the form of the trial values in Section 7. The term $(\beta^2/4)\|D - Q_N\|_F^2$ is the cost of an endpoint that differs from the prescribed overlap matrix, and it is used in Section 9.

Definition 4.16 (Constrained partition functions)**.** Let $\ell \geq 1$ and $n \geq 2$. The *overlap matrix* of $x^1, \dots, x^\ell \in \Sigma_n$ is $(R(x^a, x^b))_{a,b=1}^{\ell}$. A symmetric $\ell \times \ell$ matrix D is *feasible* if it is the overlap matrix of at least one ℓ-tuple. For feasible D, we define

$$Z_{\ell,n}(D) = \sum_{\substack{x^1, \dots, x^\ell \in \Sigma_n \\ R(x^a, x^b) = D_{ab} \ \forall a,b}} \exp\Big(\beta \sum_{a=1}^{\ell} H_n(x^a) \Big), \qquad p_{\ell,n}(D) = \frac{1}{n} \mathbb{E} \log Z_{\ell,n}(D).$$

Note that $Z_{\ell,n}(D)/Z_n^\ell$ is the probability that ℓ independent samples from π^J have overlap matrix D. Since this probability is at most one, we have $p_{\ell,n}(D) \le \ell p_n(\beta)$. Also $p_{1,n}(1) = p_n(\beta)$. A feasible matrix has unit diagonal and off-diagonal entries in $\{-1, -1+2/n, \dots, 1\}$, and there are at most $(n+1)^{\ell(\ell-1)/2}$ feasible $\ell \times \ell$ matrices. For two and three replicas we use the matrices

$$D_2(q) = \begin{pmatrix} 1 & q \\ q & 1 \end{pmatrix}, \qquad D(r,q,\delta) = \begin{pmatrix} 1 & r+\delta & r-\delta \\ r+\delta & 1 & q \\ r-\delta & q & 1 \end{pmatrix}, \tag{4.16}$$

the latter in the coordinate order (v,x,y). If $v,x,y \in \Sigma_n$ have overlap matrix $D(r,q,\delta)$, then $q = R(x,y)$, $r = (R(v,x)+R(v,y))/2$ and $\delta = (R(v,x)-R(v,y))/2$.

Remark 4.17 (Spin flips and permutations). Let ϖ be a permutation of $\{1,\dots,\ell\}$ and let $\epsilon \in \{-1,1\}^\ell$. For a symmetric D, set $(\mathsf{O}D)_{ab} = \epsilon_a \epsilon_b D_{\varpi(a)\varpi(b)}$. If $(x^b)_{b\le\ell}$ has overlap matrix D, then $(\epsilon_a x^{\varpi(a)})_{a\le\ell}$ has overlap matrix $\mathsf{O}D$, and this map is a bijection between the two sets of tuples. Since $H_n(-x) = H_n(x)$, it preserves $\sum_a H_n(x^a)$. We conclude that $\mathsf{O}D$ is feasible if and only if D is, and that $Z_{\ell,n}(\mathsf{O}D) = Z_{\ell,n}(D)$ for every realization of the disorder. In particular, $Z_{2,n}(D_2(-q)) = Z_{2,n}(D_2(q))$, reversing v gives $Z_{3,n}(D(-r,q,-\delta)) = Z_{3,n}(D(r,q,\delta))$, and exchanging x and y gives $Z_{3,n}(D(r,q,-\delta)) = Z_{3,n}(D(r,q,\delta))$.

The next two results, from [77, Section 2], are the tools used to compare constrained pressures with trial values. The first is stated for a hierarchy with a general terminal function, which is the form used in the interpolation proofs.

Proposition 4.18 (Differentiation of a hierarchy [77, Lemma 2.2])**.** *Let $\mathcal{S}$ be a nonempty finite set, let ω be a random element (the quenched randomness), let $\mathbf{g}_1,\dots,\mathbf{g}_N$ be standard Gaussian vectors, independent of each other and of ω, and let $0 \le m_1 \le \cdots \le m_N \le 1$. For $s \in \mathcal{S}$, let*

$$U(s) = a(s,\omega) + \sum_{i=1}^N \mathbf{a}_i(s)\cdot\mathbf{g}_i, \qquad \mathbb{E}\max_{s\in\mathcal{S}}|a(s,\omega)| < \infty,$$

with deterministic vectors $\mathbf{a}_i(s)$. Set $\Psi_N = \log\sum_{s\in\mathcal{S}} e^{U(s)}$, and for $j = N,\dots,1$ set $\Psi_{j-1} = m_j^{-1}\log \mathbb{E}_j e^{m_j\Psi_j}$, or $\Psi_{j-1} = \mathbb{E}_j\Psi_j$ if $m_j = 0$, where $\mathbb{E}_j$ integrates $\mathbf{g}_j$. Under the path law, given $(\omega,\mathbf{g}_1,\dots,\mathbf{g}_{j-1})$ the law of $\mathbf{g}_j$ has density $\exp\{m_j(\Psi_j - \Psi_{j-1})\}$ with respect to its Gaussian law, and the terminal configuration S has the Gibbs law proportional to $e^{U(S)}$. For $0 \le l \le N$, let $\mathbb{E}_l^{(2)}$ be the expectation under the pair law of (S^1,S^2) sharing $(\omega,\mathbf{g}_1,\dots,\mathbf{g}_l)$. Let $m_0 = 0$, $m_{N+1} = 1$, and $\Delta m_l = m_{l+1} - m_l$ for $0 \le l \le N$. Let e be a fixed vector-valued function on $\mathcal{S}$, and suppose that U depends on a parameter $v > 0$ in one of the following two ways: either $\mathbf{a}_j(s) = \sqrt{v}\,e(s)$ for some $1 \le j \le N$, or $j = 0$, $\omega = (\omega',\mathbf{g})$ with independent components, $\mathbf{g}$ a

standard Gaussian vector, and $a(s,\omega) = a_*(s,\omega') + \sqrt{v}\,\mathbf{g}\cdot e(s)$ *with* $\mathbb{E}\max_s|a_*(s,\omega')| < \infty$. *In both cases* v *occurs only in the displayed coefficient. Then, with all other parameters fixed, for* $v > 0$,

$$\frac{\partial}{\partial v}\mathbb{E}\Psi_0 = \frac{1}{2}\Big(\mathbb{E}\|e(S)\|^2 - \sum_{l=j}^{N}\Delta m_l\,\mathbb{E}_l^{(2)}e(S^1)\cdot e(S^2)\Big).$$

The identity extends to $v = 0$ *by integration and continuity.*

In [77], the quenched root is level 0 and the jth field is integrated at mass x_{j-1}; with $x_{j-1} = m_j$, the weights $x_l - x_{l-1}$ of [77] are our Δm_l. Under $\mathbb{E}_N^{(2)}$ the two copies share all fields and only the terminal draws are independent. The hypotheses on U are those of the applications in [77]: the terminal exponent is affine in the Gaussian variables, and $\mathcal{S}$ is finite, so all coefficients are bounded. Gaussian exponential integrability and the integrability of the root term then make every recursion finite and justify the differentiation, as in [77, proof of Lemma 2.2].

Used in: Section 8, in the proof of the positive fractional matrix comparison Proposition 8.3, where the first level has positive mass.

Proposition 4.19 (Matrix interpolation bound [77, Proposition 2.3])**.** *Let* $n \geq 2$ *and* $\ell \geq 1$, *let* D *be a feasible* $\ell\times\ell$ *matrix, let* (Q,m) *be an admissible path of* $\ell\times\ell$ *matrices, and let* λ *be a symmetric* $\ell\times\ell$ *matrix. Then*

$$p_{\ell,n}(D) \leq \mathcal{T}_D(Q,m,\lambda) = \Psi(Q,m,\lambda) - \lambda : D + \frac{\beta^2}{4}\Big(\|D - Q_N\|_F^2 - \int m\,\mathrm{d}\|Q\|_F^2\Big). \quad (4.17)$$

For fixed D *and* λ, *the bound passes to every trial value for* D *in the sense of Definition 4.15.*

This is [77, (2.11)], with the constrained pressure [77, (2.9)] and the path integral [77, (2.10)]. In [77] the path is indexed from $Q_{-1} = 0$; inserting the quenched level $Q_0 = 0$ allowed there and setting $x_{j-1} = m_j$ identifies the two conventions, including the penalty. The endpoint Q_N is arbitrary, subject only to positive semidefinite increments. Its diagonal need not equal one, and the cost of a mismatch with D is $(\beta^2/4)\|D - Q_N\|_F^2$. The proposition belongs to Guerra's interpolation framework [59]. As [77, Section 2] explains, Talagrand's bounds for coupled replicas [105, Propositions 3.4 and 4.4] already retain a cost for an endpoint that differs from the prescribed overlaps, and Panchenko's vector-spin formulation [89] proves the asymptotic sharpness of such bounds. The bound is quenched, and the positive fractional version needed in Section 8 is proved there.

Used in: Section 7, for the trial values of the perturbed three-replica path, with zero mismatch; Section 9, for the comparison at separated overlaps, with mismatch cost; and Section 8, which adapts it to positive fractional moments.

Lemma 4.20 (Block factorization)**.** *Let* α *be a step distribution function, and let* $0 = t_0 < t_1 < \cdots < t_p = 1$ *contain the points where* α *jumps. Write* $m_i = \alpha(t_{i-1})$ *and* $\delta_i = t_i - t_{i-1}$.

(a) (One coordinate.) *The path with $\ell = 1$, $Q_i = t_i$ and masses m_i ($1 \le i \le p$) is admissible. Its partial values for f_0 are $\Psi_i = \Phi_\alpha(t_i, \cdot)$, its trial value for $D = 1$ and $\lambda = 0$ is $\mathcal{P}_\beta(\alpha)$, and under its path law $(\mathbf{z}_1, \dots, \mathbf{z}_p)$ has the law of $(Y_{t_1}, \dots, Y_{t_p})$, where Y is the diffusion* (4.6) *of α.*

(b) (Three coordinates.) *Let $0 \le r \le q \le 1$ belong to $\{t_0, \dots, t_p\}$. In the coordinate order (v, x, y), let $\mathbf{e} = (1, 1, 1)$ and let $\mathbf{e}_v, \mathbf{e}_x, \mathbf{e}_y$ be the unit vectors. We use the following increments of Q, with their masses:*

- *common increments $\delta_i \mathbf{e}\mathbf{e}^\mathsf{T}$ at mass $m_i/3$, for $t_i \le r$;*
- *v-increments $\delta_i \mathbf{e}_v \mathbf{e}_v^\mathsf{T}$ at mass m_i, and xy-increments $\delta_i (\mathbf{e}_x + \mathbf{e}_y)(\mathbf{e}_x + \mathbf{e}_y)^\mathsf{T}$ at mass $m_i/2$, for $r < t_i \le q$;*
- *independent increments $\delta_i \mathrm{Id}$ at mass m_i, for $t_i > q$.*

Let (Q, m) take the common increments in increasing order of i, then the v- and xy-increments in any order that keeps each of these two lists in increasing order of i and has nondecreasing masses, and then the independent increments in increasing order of i. Then (Q, m) is admissible, $Q_N = D(r, q, 0)$,

$$\Psi(Q, m, 0) = 3\Phi_\alpha(0, 0), \qquad \int m \, \mathrm{d}\|Q\|_F^2 = 6 \int_0^1 t\alpha(t) \, \mathrm{d}t,$$

and $\mathcal{T}_{D(r,q,0)}(Q, m, 0) = 3\mathcal{P}_\beta(\alpha)$.

(c) (Two coordinates.) *Let $q \in \{t_0, \dots, t_p\}$. The path with increments $\delta_i (\mathbf{e}_1 + \mathbf{e}_2)(\mathbf{e}_1 + \mathbf{e}_2)^\mathsf{T}$ at mass $m_i/2$ for $t_i \le q$, followed by $\delta_i I$ at mass m_i for $t_i > q$, is admissible, its endpoint is $D_2(q)$, and its trial value for $D_2(q)$ and $\lambda = 0$ is $2\mathcal{P}_\beta(\alpha)$.*

(d) (Path law in (b).) *For $t_i \le q$, let $Y^v_{t_i}$ be the v-coordinate of the cumulative field $\mathbf{z}$ after the common increments and the v-increments with index at most i, and let $Y^x_{t_i}$ be its x-coordinate after the common increments and the xy-increments with index at most i. Under the path law of $(Q, m, 0)$, each of $(Y^v_{t_i})_{t_i \le q}$ and $(Y^x_{t_i})_{t_i \le q}$ has the law of $(Y_{t_i})_{t_i \le q}$, the two coincide for $t_i \le r$, and conditionally on $(Y^v_{t_i})_{t_i \le r}$ the families $(Y^v_{t_i})_{r < t_i \le q}$ and $(Y^x_{t_i})_{r < t_i \le q}$ are independent, each with the conditional law of $(Y_{t_i})_{r < t_i \le q}$ given $Y_r = Y^v_r$.*

Part (b) with step approximations of α_β is the base path of [77, Section 4.2] and [77, Section 7, Step 1], and its value is [77, (7.7)]. Part (d) is [77, (4.6)]. The comparisons of Part II perturb this path near r (Sections 7 and 9) or replace α by $\max\{\alpha, 3s\}$ (Section 8).

Proof. We prove (a) first and use it for the other parts. In (b), we call the common increments, the v- and xy-increments, and the independent increments the first, the middle and the last stage. The proof of (b) has three steps: admissibility and the endpoint, the value, and the penalty.

Part (a). Since α is nondecreasing and $m_p \le 1$, the path is admissible. By (4.3) at $t = t_{i-1}$, we have $\Phi_\alpha(t_{i-1},\cdot) = T_{m_i,\beta^2\delta_i}\Phi_\alpha(t_i,\cdot)$, and backward induction from $\Psi_p = \log(2\cosh\cdot) = \Phi_\alpha(1,\cdot)$ shows that $\Psi_i = \Phi_\alpha(t_i,\cdot)$ for every i. Note that the penalty is $\sum_i m_i(t_i^2 - t_{i-1}^2) = 2\int_0^1 t\alpha(t)\,\mathrm{d}t$ and that the endpoint is $1 = D$. Then the trial value is $\Phi_\alpha(0,0) - (\beta^2/2)\int_0^1 t\alpha(t)\,\mathrm{d}t = \mathcal{P}_\beta(\alpha)$. Finally, the transition of $\mathbf{z}_i$ is $\mathcal{K}_{m_i,\beta^2\delta_i,\Phi_\alpha(t_i,\cdot)}$, and Proposition 4.3(iii) states that this is the transition of Y from t_{i-1} to t_i.

Admissibility and the endpoint in part (b). Every increment is positive semidefinite. We check that the masses are nondecreasing. For $t_i \le r$ we have $t_{i-1} < r$, so that the common masses satisfy $m_i/3 \le \alpha(r)/3$. Since r and q are breakpoints, we see that the first v- and xy-masses are $\alpha(r)$ and $\alpha(r)/2$, that the first independent mass is $\alpha(q)$, and that every v- and xy-mass is at most $\alpha(q)$. Two nondecreasing lists can always be merged into one nondecreasing list that keeps the order of each, and this shows that an order as in (b) exists. The masses of (Q,m) are then nondecreasing and at most $m_p \le 1$. Summing the increments, we obtain

$$Q_N = r\,\mathbf{e}\mathbf{e}^{\mathsf{T}} + (q-r)\big(\mathbf{e}_v\mathbf{e}_v^{\mathsf{T}} + (\mathbf{e}_x+\mathbf{e}_y)(\mathbf{e}_x+\mathbf{e}_y)^{\mathsf{T}}\big) + (1-q)\mathrm{Id},$$

which has diagonal entries 1, entries r in the positions (v,x) and (v,y), and entry q in the position (x,y). This is the matrix $D(r,q,0)$.

The value in part (b). We integrate backward. The independent increments are block diagonal with three one-dimensional blocks, and f_0 is a sum of functions of single coordinates. By Lemma 4.2(c) and part (a), after the independent increments the partial value is $\sum_{a\in\{v,x,y\}}\Phi_\alpha(q,z_a)$. Let $V = \{\mathbf{z}\in\mathbb{R}^3 : z_x = z_y\}$. Note that the common, v-, and xy-increments have ranges in V and that the path starts at $0\in V$. By Lemma 4.2(e), at points of V these steps depend only on restrictions to V, which we write as functions of (z_v,z), where $z = z_x = z_y$. On V, a v-increment adds a scalar Gaussian of variance $\beta^2\delta_i$ to z_v and leaves z fixed, and an xy-increment adds a scalar Gaussian of variance $\beta^2\delta_i$ to z and leaves z_v fixed. We claim that, during the middle stage, the restriction to V of the partial value has the form

$$\Phi_\alpha(t_v,z_v) + 2\Phi_\alpha(t_{xy},z), \tag{4.18}$$

where t_v (respectively t_{xy}) is the largest $t_i \le q$ whose v-increment (respectively xy-increment) has not yet been integrated, and $t_v = r$ (respectively $t_{xy} = r$) if there is none. At the end of the middle stage this is the restriction of $\sum_{a\in\{v,x,y\}}\Phi_\alpha(q,z_a)$, with $t_v = t_{xy} = q$. If the next step backward is the v-increment with $t_i = t_v$, then Lemma 4.2(c) and (4.3) turn (4.18) into $\Phi_\alpha(t_{i-1},z_v) + 2\Phi_\alpha(t_{xy},z)$. If it is the xy-increment with $t_i = t_{xy}$, then Lemma 4.2(c) and (d), with $k = 2$, give $T_{m_i/2,\beta^2\delta_i}\big(2\Phi_\alpha(t_{xy},\cdot)\big) = 2T_{m_i,\beta^2\delta_i}\Phi_\alpha(t_{xy},\cdot) = 2\Phi_\alpha(t_{i-1},\cdot)$. This proves the claim. We conclude that, at the start of the middle stage, the restriction to V is $\Phi_\alpha(r,z_v) + 2\Phi_\alpha(r,z)$. The common increments have range spanned by

e. On this line, where $z_v = z_x = z_y = z$, the function equals $3\Phi_\alpha(r,z)$, and Lemma 4.2(d), with $k=3$, gives $T_{m_i/3,\beta^2\delta_i}\big(3\Phi_\alpha(t_i,\cdot)\big) = 3\Phi_\alpha(t_{i-1},\cdot)$ at each common increment. After all common increments, the value at 0 is $3\Phi_\alpha(0,0)$.

The penalty in part (b). During the first stage we have $Q = t\,\mathbf{e}\mathbf{e}^{\mathsf T}$ and $\|Q\|_F^2 = 9t^2$, so that this stage contributes $\sum_{t_i\le r}(m_i/3)\cdot 9(t_i^2 - t_{i-1}^2)$. During the middle stage, if the v-clock is t_v and the xy-clock is t_{xy}, then $\|Q\|_F^2 = t_v^2 + 4t_{xy}^2 + 4r^2$. A v-increment at mass m_i contributes $m_i(t_i^2 - t_{i-1}^2)$, and an xy-increment at mass $m_i/2$ contributes $(m_i/2)\cdot 4(t_i^2 - t_{i-1}^2)$. During the last stage only the diagonal changes, and each increment contributes $3m_i(t_i^2 - t_{i-1}^2)$. Every index i contributes $3m_i(t_i^2 - t_{i-1}^2)$ in total, through one increment in the first and last stages and through two increments in the middle stage. The penalty is $3\sum_{i=1}^p m_i(t_i^2 - t_{i-1}^2) = 6\int_0^1 t\alpha(t)\,\mathrm{d}t$. Since the endpoint is $D(r,q,0)$, the trial value is $3\Phi_\alpha(0,0) - (3\beta^2/2)\int_0^1 t\alpha(t)\,\mathrm{d}t = 3\mathcal{P}_\beta(\alpha)$.

Part (c). The proof is the same, with the v-coordinate removed and the common stage replaced by the xy-increments on $[0,q]$.

Part (d). Under the path law, the jth increment has density $\exp\{m_j[\Psi_j(\mathbf{z}_j) - \Psi_{j-1}(\mathbf{z}_{j-1})]\}$ with respect to its Gaussian law. By the computation of the value, we see that for a v-increment with index i the exponent is $m_i[\Phi_\alpha(t_i, Y^v_{t_i}) - \Phi_\alpha(t_{i-1}, Y^v_{t_{i-1}})]$: the term $2\Phi_\alpha(t_{xy},z)$ of (4.18) is the same before and after the step and cancels. For an xy-increment at mass $m_i/2$ the exponent is $(m_i/2)\cdot 2[\Phi_\alpha(t_i, Y^x_{t_i}) - \Phi_\alpha(t_{i-1}, Y^x_{t_{i-1}})]$, and for a common increment at mass $m_i/3$ it is $(m_i/3)\cdot 3[\Phi_\alpha(t_i,z) - \Phi_\alpha(t_{i-1},z')]$ along the diagonal. In each case the moving coordinate makes the scalar transition $\mathcal{K}_{m_i,\beta^2\delta_i,\Phi_\alpha(t_i,\cdot)}$, which depends only on its own current value, and the increments of different branches are driven by independent Gaussian variables. This proves (d), in view of part (a). □

When α is not a step function, for instance $\alpha = \alpha_\beta$, we use Lemma 4.20 with step distribution functions $\alpha_k \to \alpha$ in L^1 whose breakpoints contain r and q. By Proposition 4.3(ii), $\mathcal{P}_\beta(\alpha_k) \to \mathcal{P}_\beta(\alpha)$, and the last sentence of Proposition 4.19 applies to the limit. The path-law statement (d) also passes to the limit, in the following sense. Let $Y^{(k)}$ and Y be the scalar diffusions of α_k and α, driven by the same Brownian motion, and let $u_k = u_{\alpha_k}$. Since $|u_k| \le 1$ and u is 1-Lipschitz in space ($0 < \chi \le 1$), by Gronwall's inequality we have

$$\sup_{t\le 1}|Y^{(k)}_t - Y_t| \le \beta^2 e^{\beta^2}\big(\|\alpha_k - \alpha\|_{L^1} + \|u_k - u\|_\infty\big),$$

which tends to zero by Proposition 4.3(ii). For the two branches, we use the same Brownian path up to r and independent Brownian continuations after r. The branch laws of (d) for α_k, evaluated at any fixed finite set of common breakpoints, therefore converge to the corresponding laws of two diffusions of α that coincide up to r and are conditionally independent after r given their common value at r.

Used in: Sections 7 to 9, for the value and the law of the unperturbed path.

Proposition 4.21 (Gaussian tools [77, Lemma 2.1])**.** *Let* $\mathbf{g}$ *be a standard Gaussian vector in* $\mathbb{R}^M$*, and let* $f\colon \mathbb{R}^M \to \mathbb{R}$ *be L-Lipschitz. Then, for* $a > 0$ *and* $\theta \in \mathbb{R}$*,*

$$\mathbb{P}\big(|f(\mathbf{g}) - \mathbb{E}f(\mathbf{g})| \ge a\big) \le 2\exp\Big(-\frac{a^2}{2L^2}\Big), \qquad \log \mathbb{E}\exp\big\{\theta\big(f(\mathbf{g}) - \mathbb{E}f(\mathbf{g})\big)\big\} \le \frac{\theta^2 L^2}{2}. \tag{4.19}$$

If a probability density on $\mathbb{R}^M$ *is proportional to* e^{-V}*, where* V *is twice differentiable and* $\nabla^2 V \succeq I$*, then* $\mathrm{Var}(f) \le \mathbb{E}\|\nabla f\|^2$ *under this density for every continuously differentiable* f *with finite right side. In particular, the variance of a linear functional* $\mathbf{a}\cdot\mathbf{X}$ *of a random vector* $\mathbf{X}$ *with this density is at most* $|\mathbf{a}|^2$*.*

The bounds (4.19) are the Gaussian concentration inequality of Tsirelson, Ibragimov, and Sudakov [112], in the logarithmic-moment and tail forms of [28, Section 5.4, Theorems 5.5–5.6], and the variance bound is the case $\nabla^2 V \succeq I$ of the Brascamp–Lieb inequality [30, Theorem 4.1]. The proof in [77] covers (4.19) and the case of linear f, under an upper bound on $\nabla^2 V$ that holds in its applications. The linear case enters this paper only through Proposition 4.27.

Used in: Section 9, for the concentration of logarithms of constrained partition functions, and Section 10, in the proof of Lemma 10.1.

Remark 4.22 (Lipschitz constants of logarithms of constrained partition functions). The Gaussian vector in applications of (4.19) is $J = (J_{ij})_{i<j} \in \mathbb{R}^{n(n-1)/2}$. Let $\mathcal{S} \subset \Sigma_n^\ell$ be nonempty and $F_{\mathcal{S}}(J) = \log \sum_{(x^a)\in\mathcal{S}} \exp(\beta \sum_{a=1}^{\ell} H_n^J(x^a))$. Then $\partial F_{\mathcal{S}}/\partial J_{ij} = \beta n^{-1/2}\langle \sum_a x_i^a x_j^a \rangle_{\mathcal{S}}$, where $\langle\cdot\rangle_{\mathcal{S}}$ is the Gibbs average on $\mathcal{S}$. Each derivative has absolute value at most $\beta\ell/\sqrt{n}$. This implies that $F_{\mathcal{S}}$ is Lipschitz with constant $\beta\ell\sqrt{(n-1)/2}$; this is the bound used in [77, proof of Lemma 4.2]. It applies to $\log Z_{\ell,n}(D)$ and, with $\ell = 1$ and $\mathcal{S} = \Sigma_n$, to F_n.

4.5 Two- and three-replica pressure bounds

In this section, we assume that $\beta > 1$, we write $\alpha = \alpha_\beta$, and we let Y and Γ be the diffusion and response of α_β. The three results below are proved in [77, Sections 4 and 7] from Proposition 4.19, applied to perturbations of the paths of Lemma 4.20, and from the scalar identities of Propositions 4.3 and 4.6. We import their pressure bounds, not the probability bounds that [77] derives from them at its own scales. We convert them into probability bounds at our scales in Section 9.

Proposition 4.23 (Pair pressure bound [77, Lemma 4.3, (4.4)]). *For every $n \geq 2$ and every $q \in [0,1]$ such that $D_2(q)$ is feasible,*

$$p_{2,n}(D_2(q)) \leq 2P_\beta - \frac{1}{2}\big(\Gamma(q) - q\big)^2.$$

By Remark 4.17, the same bound holds for $q \in [-1,0]$ with $\Gamma(|q|) - |q|$ in place of $\Gamma(q) - q$. The proof in [77] uses the path of Lemma 4.20(c) and the terminal multiplier $\lambda_{12} = \lambda_{21} = L/2$ with $L = q - \Gamma(q)$. The lemma in [77] also contains a probability bound, which we do not use, for the overlaps R at which $|\Gamma(|R|) - |R||$ exceeds a fixed positive tolerance.

Used in: Section 9, for the pair overlaps at which $|q - \Gamma(q)|$ is not small and for the overlaps near one.

Proposition 4.24 (Signed fixed-scale pressure bound [77, (4.20)]). *For every $b_* \in (0,1)$, there exist $r_0, \delta_0, \nu, c_0 > 0$, depending only on β and b_*, with the following property. For every $n \geq 2$ and every feasible $D(r,q,\delta)$ with*

$$0 \leq r \leq r_0, \qquad |\delta| \leq \delta_0, \qquad b_* \leq q \leq 1, \qquad |q - \Gamma(q)| \leq \nu,$$

we have

$$p_{3,n}(D(r,q,\delta)) \leq 3P_\beta - c_0\delta^2.$$

In [77] the cutoff b_* is called h and the tolerance ν is called τ, and (4.20) is proved in [77, Sections 4.2–4.4] for nonempty constraints. No condition $r \geq |\delta|$ is imposed, and the two small overlaps $r \pm \delta$ may have opposite signs. Reversing v (Remark 4.17), we obtain the same bound for $-r_0 \leq r \leq 0$. With the conditions on q, the bound holds whenever $|r+\delta|, |r-\delta| \leq \min\{r_0, \delta_0\}$, since then $|r|$ and $|\delta|$ are at most this minimum. The perturbation in the proof acts on a fixed interval $[s_1, s_2] \subset (0, b_*)$, and the gain comes from the strict inequality $\beta^2(\int_{b_*}^1 \alpha(t)\,\mathrm{d}t)(\int_0^1 \alpha(t)\,\mathrm{d}t) < 1$, which follows from Proposition 4.6(iii). The paper [77] converts (4.20) into the probability bound of its Lemma 4.1, with tolerance $2n^{-1/100}$, using its Lemma 4.2 and its upper bound $Cn^{-1/24}$ on the mean bias. We use neither of these lemmas.

Used in: Section 9, for the fixed-scale bound (9.3), and through it for the signed overlap constraints of Sections 10 and 11.

Proposition 4.25 (Narrow-annulus comparison [77, Section 7, Steps 1–8, (7.30)–(7.31)]). *Let $t_0 > 0$ satisfy $3t_0 \leq 1$ and $\Gamma(t) = t$ for $t_0/2 < t < 3t_0$. Let $\rho_{\mathrm{OB}} = n^{-1/10000}$, $h = \rho_{\mathrm{OB}}^8$, $\varepsilon = \rho_{\mathrm{OB}}^{20}$, $\theta = \rho_{\mathrm{OB}}^2$, and $W = (2(1-\theta)^2 + \theta^2)/h$. There exist $C, c > 0$ and n_0, depending only on β and t_0, with the following properties.*

(i) *For $n \ge n_0$ and every feasible $D(r,q,\delta)$ with $t_0/2 < r < 3t_0$, $\mu_\beta([r-h,r+h]) \le \rho_{\mathrm{OB}}^4$, $q \ge r + \rho_{\mathrm{OB}}$ and $\varepsilon \le |\delta| \le 2\varepsilon$,*

$$\begin{aligned} p_{3,n}(D(r,q,\delta)) \le 3P_\beta + \delta^2\Big\{ & \beta^2 W\Big(\mathbb{E}\chi(q,Y_q) - \int_q^1 \alpha(t)\,\mathrm{d}t\Big) \\ & + 2(1-\theta)\theta\beta^2\alpha(r)\big(\beta^2\mathbb{E}[\chi(q,Y_q)\chi(r,Y_r)] - 1\big) \\ & + C(\rho_{\mathrm{OB}}^4 + \theta^2)\Big\} + C\big(|\delta|^3 h^{-3/2} + \delta^4 h^{-2}\big). \end{aligned} \tag{4.20}$$

(ii) *For every $t_0/2 < r < 3t_0$ and $r \le q \le 1$,*

$$\beta^2\mathbb{E}[\chi(q,Y_q)\chi(r,Y_r)] = 1 - \beta^4\int_r^q \alpha(t)\mathbb{E}[\chi(t,Y_t)^2\chi(r,Y_r)]\,\mathrm{d}t \le 1 - c(q-r). \tag{4.21}$$

By Corollary 4.8, the hypothesis on t_0 holds for every $t_0 \in (0, q_\beta/3]$, and finitely many intervals $(t_0/2, 3t_0)$ with such t_0 cover any compact subset of $(0, q_\beta)$. The estimate (4.20) is [77, (7.30)], obtained there from the step version [77, (7.29)] by passing to α_β at fixed n, r, q, δ. It is proved by applying Proposition 4.19 with $\lambda = 0$ to the path of Lemma 4.20(b), perturbed as follows. With the notation of Lemma 4.20(d), for step approximations whose breakpoints contain $r-h$, r, $r+h$, and q, let

$$\mathsf{Y}_0 = \frac{(Y_r^x - Y_{r-h}^x) - (Y_{r+h}^x - Y_r^x)}{h}, \qquad \mathsf{Y}_1 = \frac{Y_{r+h}^v - Y_r^v}{h}, \qquad \mathsf{Y} = (1-\theta)\mathsf{Y}_0 + \theta\mathsf{Y}_1,$$

as linear combinations of the Gaussian increments [77, (7.8)]. The x field receives $+\delta$ times, and the y field $-\delta$ times, the contribution of each increment to Y. The endpoint becomes $D(r,q,\delta) + \delta^2 W(\mathbf{e}_x - \mathbf{e}_y)(\mathbf{e}_x - \mathbf{e}_y)^{\mathsf{T}}$, and the mismatch cost $\beta^2\delta^4 W^2$ [77, (7.13)] is part of the last error term. The coefficient W is denoted K in [77]. The identity in (4.21) follows from Proposition 4.3(iv) and $\Gamma'(r) = 1$, and the inequality is [77, (7.31)].

Proposition 4.25 is a consequence of the proof in [77, Section 7] rather than a verbatim statement there, and we record how it is obtained. The setup of [77, Section 7] also imposes $3t_0 < q_*$, with q_* the constant of Proposition 4.29, and, before Step 1, the restriction $|q - \Gamma(q)| \le \rho_{\mathrm{OB}}^{13}$. The Gaussian calculations through (7.30) use neither restriction. Their geometric requirements are $0 < r - h$ and $r + h < q \le 1$, which hold under the hypotheses of (i) for all sufficiently large n, and the pair restriction is needed only to obtain (7.32). In [77, Section 7], (7.31) is stated for the pairs (r,q) of Lemma 7.1, but its proof uses only $t_0/2 < r < 3t_0$ and $r \le q \le 1$. Indeed, let M be such that the event $\{\beta \sup_{t\le 1}|B_t| + \beta^2 \le M\}$ has probability $p_M > 0$; on it $|Y_t| \le M$ for all $t \le 1$. With $c_M = \min_{0\le t\le 1,\,|z|\le M}\chi(t,z) > 0$, we have $\mathbb{E}[\chi(t,Y_t)^2\chi(r,Y_r)] \ge p_M c_M^3$, and $\alpha(t) \ge \alpha(t_0/2) > 0$ for $t \ge r$. We may take $c = \beta^4\alpha(t_0/2)p_M c_M^3$. After imposing the additional restriction $|q - \Gamma(q)| \le \rho_{\mathrm{OB}}^{13}$, [77] deduces from (4.20) and (4.21) the deficit $c\delta^2\rho_{\mathrm{OB}}^3$ [77, (7.32)] and the probability bound of its Lemma 7.1. We use neither.

Used in: Section 9, for the comparison at separated overlaps in (9.17) and (9.18).

Remark 4.26 (Where the scales enter the proof of (4.20)). Section 9 needs a version of (4.20) with h and θ fixed, and with the oscillation $\alpha(r+h)-\alpha(r-h)$ in place of ρ_{OB}^4. This version is not stated in [77], and we derive it in Section 9 by following Steps 1–7 of [77, Section 7]. In those steps, the specific scales enter only at the following places.

(a) Step 1 uses $0 < r-h$ and $r+h < q$, so that the triangular function on $[r-h, r+h]$ that defines the triangular covariance perturbation vanishes at 0 and at q.

(b) The errors $O(\rho_{\mathrm{OB}}^4)$ in [77, (7.14), (7.15), (7.20), (7.23)–(7.28)] come from the oscillation bound $\alpha(r+h)-\alpha(r-h) \le \rho_{\mathrm{OB}}^4$ [77, (7.6)], which follows from the hypothesis on $\mu_\beta([r-h, r+h])$. Each of them is at most a constant times $\alpha(r+h)-\alpha(r-h)$. The ramp error is $O(1)$ and is multiplied by θ^2.

(c) Step 4 uses $\delta^2/h \to 0$ only to bound $\mathbb{E}[\mathsf{Y}^4 e^{\delta^2 \mathsf{Y}^2}]$ through Gaussian exponential moments, since Y is a centered Gaussian of variance $\beta^2 W$ plus a term bounded by a constant depending on β; a bound $\delta^2/h \le c_\beta$ with $c_\beta > 0$ small, together with $h \le 1$, suffices.

(d) Once the geometric requirements in (a) and the moment restriction in (c) are satisfied, the quantitative lower bound $q - r \ge \rho_{\mathrm{OB}}$, the restriction $|q - \Gamma(q)| \le \rho_{\mathrm{OB}}^{13}$, and the relations between the powers of ρ_{OB} that define h, ε and θ are used only in Steps 8 and 9.

At fixed h and θ, the last error term of (4.20) is absorbed only for small amplitudes, so the fixed-parameter version in Section 9 holds under a condition $|\delta| \le \delta_0$. It must be derived from Steps 1–7 with the oscillation of α kept explicit; it does not follow by applying Proposition 4.25(i) at fixed h and θ.

4.6 Energy, disorder rotation, and overlap-neighborhood masses

In this section, we assume that $\beta > 1$ and that $J, \widetilde{J}$ are independent standard Gaussian vectors in $\mathbb{R}^{n(n-1)/2}$, indexed by pairs $i < j$. We write Z_n^J, π^J, and H_n^J to indicate the disorder. The three results below, from [77, Sections 5–6], use the off-diagonal Hamiltonian without the auxiliary diagonal variables.

Proposition 4.27 (Upper energy tail [77, Lemma 5.1])**.** *There exist constants $e_1 \in (0, \beta/2)$ and $c_E > 0$ such that, for all sufficiently large n,*

$$\mathbb{P}_J\Big(\pi^J\{x \in \Sigma_n : H_n^J(x) > n e_1\} > e^{-c_E n}\Big) \le e^{-c_E n}.$$

The proof in [77] combines the strict bound $n^{-1}\mathbb{E}\pi^J(H_n^J) \le e_0 < \beta/2$ for all sufficiently large n of [77, Lemma 3.3], which uses $\beta > 1$, with the linear case of the variance bound in Proposition 4.21 for the law of J given a Gibbs sample.

Used in: Section 10, in the proof of Lemma 10.1.

Proposition 4.28 (Disorder rotation [77, Lemma 5.2 and (5.5)])**.** *For $\theta \in (0, \pi/2)$, let $J_\theta = \cos\theta\, J + \sin\theta\, \widetilde{J}$. For every $n \geq 2$, every $s \in \mathbb{R}$ and every $u > 0$,*

$$\mathbb{E}\exp\{s(\log Z_n^{J_\theta} - \log Z_n^J)\} \leq \exp\Big\{\frac{s^2\theta^2\beta^2(n-1)}{4}\Big\}$$

and

$$\mathbb{P}\big(\log Z_n^{J_\theta} - \log Z_n^J \geq u\big) \leq \exp\Big\{-\frac{u^2}{\theta^2\beta^2(n-1)}\Big\}.$$

If $0 < t < 1/2$ and $J' = \sqrt{1-t}\, J + \sqrt{t}\, \widetilde{J}$, then for every $u > 0$,

$$\mathbb{P}\big(\log Z_n^{J'} - \log Z_n^J \geq u\big) \leq \exp\Big(-\frac{u^2}{2\beta^2 n t}\Big).$$

The first two bounds are [77, (5.3)–(5.4)]. The last is [77, (5.5)], which follows from the second with $\theta = \arcsin\sqrt{t}$, since $\theta^2 \leq 2t$ for $0 < t < 1/2$. All three bounds are uniform in θ and t. They also apply to rotations whose size depends on n, and they concern the joint law of $(J, \widetilde{J})$.

Used in: Section 10, in the proof of Lemma 10.1.

Proposition 4.29 (Overlap-neighborhood mass bound [77, Proposition 6.1])**.** *Let $L_{\mathrm{OB}} = n^{1/10000}$. There exists $q_* \in (0,1)$, depending only on β, with the following property. For every fixed $D_1 > 0$ there exist $D_3 = D_3(\beta, D_1) > 0$ and events $\mathcal{C}_n(D_1)$, measurable with respect to J, such that $\mathbb{P}_J(\mathcal{C}_n(D_1)) \to 1$ and, on $\mathcal{C}_n(D_1)$, for every $S \subset \Sigma_n$,*

$$\pi^J\Big(\Big\{x \in S : \pi^J\{z \in S : |R(x,z)| \geq q_*\} < e^{-D_3 L_{\mathrm{OB}}}\Big\}\Big) < e^{-D_1 L_{\mathrm{OB}}}.$$

Note that both masses in the display are masses under π^J itself, not under π^J conditioned on S. The proposition, titled "Stretched-exponential coverage" in [77], is stated there at the scale L_{OB} only. Its proof uses this scale only through $L_{\mathrm{OB}} \to \infty$ and $L_{\mathrm{OB}}/n \to 0$, and q_* is chosen before D_1 from the constant e_1 of Proposition 4.27. Lemma 10.1 in Section 10 states and proves the corresponding bound for every deterministic sequence $L_n \to \infty$ with $L_n = o(n)$, by the argument of [77, Section 6], using Propositions 4.21, 4.27, and 4.28.

Used in: Section 10, as the statement extended in Lemma 10.1.

PART II

Static estimates for the lower bound

SECTION 5

The scalar Parisi problem on the support

In this section, we collect the properties of the scalar Parisi equation used in the barrier argument. Recall from Section 4 the solution Φ_α of the Parisi equation for a distribution function α, its spatial derivatives u, χ, and w, the scalar diffusion Y_t, and the function $\Gamma(t) = \mathbb{E}u(t, Y_t)^2$. We first prove two statements that hold for every distribution function. In Lemma 5.1, we bound each positive spatial derivative of Φ_α by a constant that depends only on β and on the order of the derivative. In Lemma 5.2, we record the stochastic differential equations that these derivatives satisfy along Y. We then specialize to the Parisi distribution function α_β on the interval $(0, q_\beta)$, where $\Gamma(t) = t$. In Lemma 5.3, we prove the two stationarity identities $\beta^2\mathbb{E}\chi(t, Y_t)^2 = 1$ and $\mathbb{E}w(t, Y_t)^2 = 2\alpha(t)\mathbb{E}\chi(t, Y_t)^3$, the formula $\alpha = \mathbb{E}w^2/(2\mathbb{E}\chi^3)$, smoothness and a Lipschitz bound for α, and bounds on mixed derivatives of Φ on compact subintervals. The positive constant m_β of (5.20) is defined in Corollary 5.5.

We use three external inputs: Proposition 4.3, on the regularity and stability of the scalar equation; Theorem 4.7, which states that the support of μ_β is $[0, q_\beta]$; and the identity (4.10), which states that $\Gamma(t) = t$ for $0 < t < q_\beta$. The derivative bounds are used in Sections 6 to 8. The stationarity identities enter the coefficient computation of Section 7, and the constant m_β enters the cubic pressure deficit there and, through it, the choice of parameters in Sections 8, 10, and 12.

The statements of this section are known, some of them in a slightly different form. We include proofs because the later sections need them with explicit uniformity. For each fixed distribution function, bounds on all positive spatial derivatives, and their convergence when the profile is approximated, were proved by Auffinger and Chen [12, Propositions 1–2] (see also [69, Proposition 2.4]). The bounds of order at most three, the martingale identity (5.9), and the expectation of (5.10) are proved in [77, Lemma 3.1]. Our proof of Lemma 5.1 extends the Feynman–Kac argument given there to every order. The identities (5.8) with $j \le 3$ and the first two identities of (5.12) appear in [69, Lemma 2.5, (2.12)–(2.16)]; the formulas for Γ' given there are stochastic versions of the differentiation formulas of [12, Proposition 3]. The identity $\Gamma(q) = q$ at points of the support goes back to Talagrand [104] (see [12, Theorem 5 and its proof]); it is part of the characterization of Parisi measures by Chen [36, Proposition 1] and by Jagannath and Tobasco [63, Proposition 1.1], and we use it in the form proved in [77, Proposition 3.2]. Since $\operatorname{supp}\mu_\beta = [0, q_\beta]$ [69, Theorem 1.1], the identity holds on all of $(0, q_\beta)$. In the

proof of [12, Theorem 2(ii)], Auffinger and Chen derive the quotient formula for α and use an induction to prove its smoothness. The same quotient argument, including the passage from almost every point to every point, was carried out near $t = 0$ in [8, proof of Lemma 7.2, Step 3, (7.8)–(7.12)], in a rescaled normalization. The smoothness of α_β on $(0, q_\beta)$ is also part of Theorem 4.7 (see Corollary 4.8). The use of the second stationarity identity in the covariance expansion of Section 7 is new.

Notation. Throughout this section, $\beta > 1$ is fixed and $b = \beta^2$. We use the conventions of Section 4.2. A distribution function is a nondecreasing, right-continuous function $\alpha\colon [0,1] \to [0,1]$ with $\alpha(1) = 1$; equivalently, $\alpha(t) = \mu([0,t])$ for a probability measure μ on $[0,1]$. Let α be a step distribution function, with breakpoints $0 = t_0 < t_1 < \cdots < t_p = 1$ and $\alpha = m_i$ on $[t_{i-1}, t_i)$. Then $\Phi_\alpha(1,z) = \log(2\cosh z)$, and the Gaussian recursion (4.3) reads

$$\Phi_\alpha(t,z) = \frac{1}{m_i} \log \mathbb{E} \exp\Big(m_i\, \Phi_\alpha\big(t_i, z + \beta\sqrt{t_i - t}\, Z\big)\Big), \qquad t_{i-1} \le t < t_i, \tag{5.1}$$

where Z is a standard Gaussian variable, and where the right side is read as $\mathbb{E}\Phi_\alpha(t_i, z + \beta\sqrt{t_i - t}\, Z)$ when $m_i = 0$. For a general distribution function α, the function Φ_α is the uniform limit of Φ_{α_N} along any sequence of step distribution functions with $\|\alpha_N - \alpha\|_{L^1} \to 0$ (Proposition 4.3 (i)). We write

$$f_0 = \Phi_\alpha, \qquad f_j = f_j^\alpha = \partial_z^j \Phi_\alpha \quad (j \ge 1),$$

so that $f_1 = u$, $f_2 = \chi$, and $f_3 = w$. For $(t,z) \in [0,1] \times \mathbb{R}$, let $(Y_r^{t,z})_{t \le r \le 1}$ be the solution of

$$\mathrm{d}Y_r^{t,z} = \beta\, \mathrm{d}B_r + b\, \alpha(r)\, u(r, Y_r^{t,z})\, \mathrm{d}r, \qquad Y_t^{t,z} = z,$$

where B is a standard Brownian motion. This equation has a unique strong solution (Proposition 4.3 (iii) and (v)), and the scalar diffusion of Section 4 is $Y = Y^{0,0}$. We write $\|\cdot\|_\infty$ for the supremum over $[0,1] \times \mathbb{R}$, or over $\mathbb{R}$ for a function of z alone. From Proposition 4.3, we use the bounds $|u| \le 1$ and $0 < \chi \le 1$ (the proof of [77, Lemma 3.1] gives the sharper bound $\chi \le 1 - u^2$, which we do not need), the joint continuity of Φ_α, u, and χ, the well-posedness of the scalar diffusion, and the stability estimate in part (ii) there,

$$\sup_{t,z} \big(|\Phi_\alpha - \Phi_{\tilde\alpha}| + |u_\alpha - u_{\tilde\alpha}| + |\chi_\alpha - \chi_{\tilde\alpha}| \big) \le C_\beta \|\alpha - \tilde\alpha\|_{L^1},$$

which holds for all distribution functions α and $\tilde\alpha$.

5.1 Bounds on spatial derivatives

The derivative bounds below are uniform over all distribution functions. We need this uniformity because the later sections apply the bounds to step approximations and to

profiles other than α_β, such as $\max\{\alpha_\beta, 3s\}$ with $0 \le s \le 1/3$. The constants are defined by $C_1 = C_2 = 1$ and, for $j \ge 3$,

$$C_j = e^{jb}\Big(\|\partial_z^j \log(2\cosh z)\|_\infty + \frac{b}{2} \sum_{a=2}^{j-2} \binom{j}{a} C_{a+1} C_{j+1-a} \Big). \tag{5.2}$$

The sum is empty for $j = 3$, and it involves only $C_3, \dots, C_{j-1}$. Every derivative of positive order of $\log(2\cosh z)$ is a polynomial in $\tanh z$ and is bounded.

Lemma 5.1 (Bounds on spatial derivatives)**.** *Let α be a distribution function. For all $j \ge 1$, the derivative $f_j = \partial_z^j \Phi_\alpha$ exists and is continuous on $[0,1] \times \mathbb{R}$, and*

$$\sup_{(t,z) \in [0,1] \times \mathbb{R}} |\partial_z^j \Phi_\alpha(t,z)| \le C_j. \tag{5.3}$$

Moreover, the following hold.

(i) *If α_N are distribution functions such that $\|\alpha_N - \alpha\|_{L^1} \to 0$, then $\|\partial_z^j \Phi_{\alpha_N} - \partial_z^j \Phi_\alpha\|_\infty \to 0$ for all $j \ge 1$.*

(ii) *For all $j \ge 0$, $t \in [0,1]$, and $z \in \mathbb{R}$,*

$$f_j(t,z) = \partial_z^j \log(2\cosh z) + \frac{b}{2} \int_t^1 \big(f_{j+2} + \alpha\, \partial_z^j(u^2) \big)(r,z)\, \mathrm{d}r, \tag{5.4}$$

where $\partial_z^j(u^2) = \sum_{a=0}^{j} \binom{j}{a} f_{a+1} f_{j+1-a}$.

The bounds concern derivatives of positive order only; the function Φ_α itself grows linearly in z. By (5.4), the Parisi equation and its spatial derivatives hold for almost every t, and at every t at which α is continuous.

Proof. We first recall why, for a fixed step distribution function, all spatial derivatives exist and are bounded, with bounds that may depend on the steps. We then write the equation for f_j as a linear equation and use its Feynman–Kac representation ([64, Theorem 5.7.6], after an exponential rescaling for a bounded potential) to obtain bounds independent of the steps, extending the third-derivative argument in [77, proof of Lemma 3.1]. Finally, we pass to general distribution functions by a forward difference argument and prove (5.4).

Step 1: a fixed step distribution function. Let α be a step distribution function as in (5.1), fix i, and let $\phi = \Phi_\alpha(t_i, \cdot)$. Suppose that ϕ is smooth, that all its derivatives of positive order are bounded, and that $|\phi(z)| \le c(1 + |z|)$ for some c. These properties hold for $\phi(z) = \log(2\cosh z)$, which is the case $i = p$. For $t_{i-1} \le t \le t_i$, the function $\Phi_\alpha(t, \cdot)$ is $1/m_i$ times the logarithm of the convolution of $e^{m_i \phi}$ with the centered Gaussian density of variance $b(t_i - t)$, or the convolution of ϕ itself when $m_i = 0$. The linear

growth of ϕ justifies differentiation under the integral. We find that, for $j \geq 1$, $f_j(t,z)$ is a polynomial, with coefficients depending only on j and m_i, in expectations of products of $\phi'(z+\sigma Z), \dots, \phi^{(j)}(z+\sigma Z)$ under the probability measure with density proportional to $e^{m_i \phi(z+\sigma Z)}$, where $\sigma = \beta\sqrt{t_i - t}$. In particular, f_j is bounded and continuous on $[t_{i-1}, t_i] \times \mathbb{R}$. For $t < t_i$, the Gaussian convolution is smooth in (t,z), and $e^{m_i \Phi_\alpha}$ (or Φ_α itself, when $m_i = 0$) solves the backward heat equation $-\partial_t v = (b/2)\partial_{zz} v$. It follows that Φ_α is smooth on $[t_{i-1}, t_i) \times \mathbb{R}$, where it satisfies

$$-\partial_t \Phi_\alpha = \frac{b}{2}\big(\partial_{zz}\Phi_\alpha + m_i(\partial_z \Phi_\alpha)^2\big).$$

The function $\Phi_\alpha(t_{i-1}, \cdot)$ has the three properties assumed for ϕ (it grows at most linearly because $|u| \leq 1$). Proceeding backward through the steps, we obtain these properties on every strip; this argument is recorded in the proof of [77, Lemma 3.1]. The bounds it gives depend on the number and the location of the steps, and the next two steps remove this dependence.

Step 2: the differentiated equation. Differentiating the equation j times in z on each strip $[t_{i-1}, t_i) \times \mathbb{R}$, we find that

$$-\partial_t f_j = \frac{b}{2}\big(f_{j+2} + \alpha\, \partial_z^j(u^2)\big) \qquad (j \geq 0), \tag{5.5}$$

with $\partial_z^j(u^2)$ as in (5.4). Let $j \geq 3$. In the sum defining $\partial_z^j(u^2)$, the terms with $a \in \{0, j\}$ add up to $2u f_{j+1}$, and the terms with $a \in \{1, j-1\}$ add up to $2j\chi f_j$. Then (5.5) becomes

$$-\partial_t f_j = \frac{b}{2}\partial_{zz} f_j + b\alpha u\, \partial_z f_j + jb\alpha\chi f_j + \mathcal{Q}_j, \qquad \mathcal{Q}_j = \frac{b\alpha}{2}\sum_{a=2}^{j-2}\binom{j}{a} f_{a+1} f_{j+1-a}. \tag{5.6}$$

The sum defining $\mathcal{Q}_j$ is empty for $j = 3$, and its factors are derivatives of orders between 3 and $j-1$. We regard (5.6) as a linear equation for f_j, with coefficients u and χ and source term $\mathcal{Q}_j$.

Step 3: bounds independent of the steps. We prove (5.3) for step distribution functions by induction on j. For $j \in \{1,2\}$, it is the bound $|u| \leq 1$, $0 < \chi \leq 1$ of Proposition 4.3. Let $j \geq 3$, and suppose that $\|f_l\|_\infty \leq C_l$ for $1 \leq l \leq j-1$. Since $0 \leq \alpha \leq 1$, we have

$$\|\mathcal{Q}_j\|_\infty \leq \frac{b}{2}\sum_{a=2}^{j-2}\binom{j}{a} C_{a+1} C_{j+1-a}.$$

Fix $(t,z) \in [0,1) \times \mathbb{R}$ and let $Y_r = Y_r^{t,z}$. We abbreviate

$$\Xi_r = jb\int_t^r \alpha(r')\chi(r', Y_{r'})\,\mathrm{d}r', \qquad t \leq r \leq 1,$$

so that $0 \le \Xi_r \le jb$. By (5.6), the drift of $f_j(r, Y_r)$ in Itô's formula is

$$\partial_r f_j + b\alpha u\, \partial_z f_j + \frac{b}{2}\partial_{zz} f_j = -jb\alpha\chi f_j - \mathcal{Q}_j,$$

evaluated at (r, Y_r). On each interval $[t_{i-1}, t_i) \cap [t, 1]$, we then have

$$\mathrm{d}\big(e^{\Xi_r} f_j(r, Y_r)\big) = -e^{\Xi_r}\mathcal{Q}_j(r, Y_r)\,\mathrm{d}r + \beta e^{\Xi_r} f_{j+1}(r, Y_r)\,\mathrm{d}B_r.$$

Recall from Step 1 that the functions f_j and f_{j+1} are bounded and continuous on $[0,1] \times \mathbb{R}$. The process

$$e^{\Xi_r} f_j(r, Y_r) + \int_t^r e^{\Xi_{r'}}\mathcal{Q}_j(r', Y_{r'})\,\mathrm{d}r', \qquad t \le r \le 1,$$

is therefore continuous, and it is a martingale. Comparing its expectations at $r = t$ and $r = 1$, we obtain

$$f_j(t,z) = \mathbb{E}\Big[e^{\Xi_1}\,\partial_z^j \log(2\cosh z)\big|_{z=Y_1} + \int_t^1 e^{\Xi_r}\mathcal{Q}_j(r, Y_r)\,\mathrm{d}r\Big].$$

Since $0 \le \Xi_r \le jb$ for $t \le r \le 1$, we have

$$|f_j(t,z)| \le e^{jb}\big(\|\partial_z^j \log(2\cosh z)\|_\infty + \|\mathcal{Q}_j\|_\infty\big) \le C_j.$$

The same bound holds at $t = 1$, which completes the induction. The elapsed time is at most one, and the only property of α we used is $0 \le \alpha \le 1$, so the constants depend neither on the number nor on the location of the steps.

Step 4: general distribution functions. Let α be a distribution function, and let α_N be step distribution functions such that $\|\alpha_N - \alpha\|_{L^1} \to 0$; for instance, $\alpha_N(t) = \alpha((k+1)/N)$ for $k/N \le t < (k+1)/N$ and $\alpha_N(1) = 1$, which converges to α pointwise by right continuity. We write $f_j^N = f_j^{\alpha_N}$. The stability estimate of Proposition 4.3 implies that Φ_{α_N}, f_1^N, and f_2^N converge uniformly to Φ_α, u, and χ. Let $j \ge 3$, and suppose that f_{j-1}^N converges uniformly to $f_{j-1} = \partial_z^{j-1}\Phi_\alpha$, which is continuous. For $\zeta > 0$, Taylor's theorem and Step 3 imply that

$$\Big|f_j^N(t,z) - \frac{f_{j-1}^N(t, z+\zeta) - f_{j-1}^N(t,z)}{\zeta}\Big| \le \frac{C_{j+1}\zeta}{2},$$

and that

$$\|f_j^N - f_j^M\|_\infty \le \frac{2}{\zeta}\|f_{j-1}^N - f_{j-1}^M\|_\infty + C_{j+1}\zeta.$$

Given $\varepsilon > 0$, we first choose ζ with $C_{j+1}\zeta < \varepsilon/2$, and then N_0 such that the first term is less than $\varepsilon/2$ for $N, M \ge N_0$. Then f_j^N converges uniformly to a continuous function $\widehat{f}_j$

with $\|\widehat{f_j}\|_\infty \le C_j$. Letting $N \to \infty$ in $f_{j-1}^N(t,z) = f_{j-1}^N(t,0) + \int_0^z f_j^N(t,y)\,\mathrm{d}y$, we find that $\partial_z f_{j-1}$ exists and equals $\widehat{f_j}$. By induction on j, this proves the existence and continuity of f_j, the bound (5.3), and the uniform convergence $f_j^N \to f_j$ for all $j \ge 1$. Now that (5.3) holds for all distribution functions, the same argument applies to any sequence of distribution functions α_N with $\|\alpha_N - \alpha\|_{L^1} \to 0$, step functions or not. This proves (i).

Step 5: the integrated equation. Let α be a step distribution function and $j \ge 0$. For fixed z, the function $t \mapsto f_j(t,z)$ is continuous on $[0,1]$ by Step 1, and continuously differentiable on each interval $[t_{i-1}, t_i)$, with derivative given by (5.5). Integrating over the steps, we obtain (5.4). For a general α, we choose α_N as in Step 4 and write $u_N = f_1^N$. Then $\Phi_{\alpha_N} \to \Phi_\alpha$ and $f_l^N \to f_l$ uniformly for all $l \ge 1$, and

$$\left| \int_t^1 \left(\alpha_N \partial_z^j(u_N^2) - \alpha\, \partial_z^j(u^2)\right)(r,z)\,\mathrm{d}r \right| \le \|\alpha_N - \alpha\|_{L^1} \|\partial_z^j(u_N^2)\|_\infty + \|\partial_z^j(u_N^2) - \partial_z^j(u^2)\|_\infty.$$

The right side tends to zero by (5.3) and the uniform convergence. Letting $N \to \infty$ in (5.4) for α_N yields (5.4) for α, which completes the proof. □

5.2 Stochastic identities along the scalar diffusion

We now record the equations satisfied by the spatial derivatives of Φ_α along the scalar diffusion. As usual, a differential identity such as (5.8) below means that the corresponding integrated identity holds almost surely, simultaneously for all $0 \le s \le t \le 1$. For a polynomial Θ in the variables $x_1, \dots, x_k$, we define the polynomials in $x_1, \dots, x_{k+1}$

$$\Theta^0 = \frac{b}{2} \sum_{i,l=1}^{k} x_{i+1} x_{l+1}\, \partial_i \partial_l \Theta, \qquad \Theta^1 = -b \sum_{i=1}^{k} \partial_i \Theta \sum_{a=1}^{i-1} \binom{i-1}{a} x_{a+1} x_{i+1-a}, \tag{5.7}$$

where ∂_i is the derivative in x_i. We write $\Theta_t = \Theta(f_1(t,Y_t), \dots, f_k(t,Y_t))$, and Θ_t^0 and Θ_t^1 for the analogous expressions with $k+1$ arguments.

Lemma 5.2 (Stochastic identities along the scalar diffusion)**.** *Let α be a distribution function, and let $Y = Y^{0,0}$. Write $u_t = u(t,Y_t)$, $\chi_t = \chi(t,Y_t)$, and $w_t = w(t,Y_t)$.*

(i) *For all $j \ge 1$,*

$$\mathrm{d}f_j(t,Y_t) = \beta f_{j+1}(t,Y_t)\,\mathrm{d}B_t - b\,\alpha(t) \sum_{a=1}^{j-1} \binom{j-1}{a} f_{a+1}(t,Y_t) f_{j+1-a}(t,Y_t)\,\mathrm{d}t. \tag{5.8}$$

In particular,

$$\mathrm{d}u_t = \beta \chi_t\,\mathrm{d}B_t, \tag{5.9}$$

$$\mathrm{d}\chi_t = \beta w_t\,\mathrm{d}B_t - b\,\alpha(t)\chi_t^2\,\mathrm{d}t. \tag{5.10}$$

(ii) *Let Θ be a polynomial in $x_1, \dots, x_k$. Then $t \mapsto \mathbb{E}\Theta_t$ is Lipschitz on $[0,1]$, and for $0 \le s \le t \le 1$,*

$$\mathbb{E}\Theta_t - \mathbb{E}\Theta_s = \int_s^t \big(\mathbb{E}\Theta_r^0 + \alpha(r)\mathbb{E}\Theta_r^1\big)\,\mathrm{d}r. \tag{5.11}$$

(iii) *For $0 \le s \le t \le 1$,*

$$\begin{aligned}
\Gamma(t) - \Gamma(s) &= b\int_s^t \mathbb{E}\chi_r^2\,\mathrm{d}r,\\
\mathbb{E}\chi_t^2 - \mathbb{E}\chi_s^2 &= b\int_s^t \big(\mathbb{E}w_r^2 - 2\alpha(r)\mathbb{E}\chi_r^3\big)\,\mathrm{d}r,\\
\mathbb{E}w_t^2 - \mathbb{E}w_s^2 &= b\int_s^t \mathbb{E}\big(f_4(r,Y_r)^2 - 6\alpha(r)\chi_r w_r^2\big)\,\mathrm{d}r,\\
\mathbb{E}\chi_t^3 - \mathbb{E}\chi_s^3 &= 3b\int_s^t \mathbb{E}\big(\chi_r w_r^2 - \alpha(r)\chi_r^4\big)\,\mathrm{d}r.
\end{aligned} \tag{5.12}$$

In particular, Γ is continuously differentiable with $\Gamma'(t) = b\mathbb{E}\chi_t^2$. The function $t \mapsto \mathbb{E}\chi_t^2$ has a right derivative at every $t \in [0,1)$, and a derivative at every point of continuity of α, given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\chi_t^2 = b\big(\mathbb{E}w_t^2 - 2\alpha(t)\mathbb{E}\chi_t^3\big). \tag{5.13}$$

Parts (i) and (ii) hold, with the same proof, for the diffusion $Y^{t,z}$ started from any $(t,z) \in [0,1]\times\mathbb{R}$, on the time interval $[t,1]$.

Proof. We prove (i) first for step distribution functions, for which the equation holds classically between the breakpoints, and then pass to the limit along a coupling of the diffusions. Parts (ii) and (iii) then follow from Itô's formula [64, Theorem 3.3.6].

Step 1: step distribution functions. Let α be a step distribution function. By Step 1 of the proof of Lemma 5.1, each f_j is smooth on the strips $[t_{i-1},t_i)\times\mathbb{R}$, bounded, and continuous on $[0,1]\times\mathbb{R}$. On each strip, Itô's formula shows that the martingale part of $f_j(t,Y_t)$ is $\beta f_{j+1}(t,Y_t)\,\mathrm{d}B_t$. Using (5.5), we find that its drift is

$$\partial_t f_j + b\alpha u f_{j+1} + \frac{b}{2}f_{j+2} = -\frac{b\alpha}{2}\sum_{a=0}^{j}\binom{j}{a}f_{a+1}f_{j+1-a} + b\alpha u f_{j+1} = -\frac{b\alpha}{2}\sum_{a=1}^{j-1}\binom{j}{a}f_{a+1}f_{j+1-a},$$

evaluated at (t,Y_t). In the second equality, we used that the terms with $a \in \{0,j\}$ equal $2uf_{j+1}$. By Pascal's rule $\binom{j}{a} = \binom{j-1}{a} + \binom{j-1}{a-1}$ and the substitution $a \mapsto j-a$ in the sum containing $\binom{j-1}{a-1}$, we have

$$\sum_{a=1}^{j-1}\binom{j}{a}f_{a+1}f_{j+1-a} = 2\sum_{a=1}^{j-1}\binom{j-1}{a}f_{a+1}f_{j+1-a}.$$

This proves (5.8) on each interval $[t_{i-1},t_i)$. The process $f_j(t,Y_t)$ is continuous in t, and the integrated identities on the successive intervals add up to (5.8) on $[0,1]$.

Step 2: general distribution functions. Let α be a distribution function, let α_N be step distribution functions such that $\|\alpha_N - \alpha\|_{L^1} \to 0$, and let Y^N be the scalar diffusion of α_N, driven by the same Brownian motion B as Y. Let $u_N = f_1^{\alpha_N}$. Recall that $|u_N| \le 1$, and that u is 1-Lipschitz in z because $|\chi| \le 1$. Then

$$|Y_t^N - Y_t| \le b \int_0^t |\alpha_N(r) - \alpha(r)| \, \mathrm{d}r + b\|u_N - u\|_\infty + b \int_0^t |Y_r^N - Y_r| \, \mathrm{d}r,$$

and Gronwall's inequality yields

$$\sup_{0 \le t \le 1} |Y_t^N - Y_t| \le be^b \big(\|\alpha_N - \alpha\|_{L^1} + \|u_N - u\|_\infty \big)$$

for every realization of B. The right side tends to zero by the stability estimate of Proposition 4.3. Using Lemma 5.1, we find that, for each $l \ge 1$,

$$|f_l^{\alpha_N}(t, Y_t^N) - f_l(t, Y_t)| \le \|f_l^{\alpha_N} - f_l\|_\infty + C_{l+1} |Y_t^N - Y_t|,$$

which tends to zero uniformly in t and in the realization. We now let $N \to \infty$ in (5.8) for α_N, in its integrated form between fixed times $s \le t$. The terms $f_j(t, Y_t)$ and $f_j(s, Y_s)$ converge uniformly. The drift integrals also converge uniformly, because their integrands are bounded by Lemma 5.1 and converge uniformly, and $\|\alpha_N - \alpha\|_{L^1} \to 0$. By Itô's isometry, the stochastic integrals converge in L^2. We conclude that (5.8) holds almost surely for each fixed $s \le t$. Since both sides are continuous in (s, t), it holds almost surely for all $s \le t$ simultaneously. The cases $j = 1$ and $j = 2$ are (5.9) and (5.10).

Step 3: polynomial moments. It follows from (i) that the processes $f_1(t, Y_t), \dots, f_{k+1}(t, Y_t)$ are continuous semimartingales with $\mathrm{d}\langle f_i(\cdot, Y), f_l(\cdot, Y)\rangle_t = b\, f_{i+1}(t, Y_t) f_{l+1}(t, Y_t) \, \mathrm{d}t$. By Itô's formula, we have

$$\mathrm{d}\Theta_t = \big(\Theta_t^0 + \alpha(t)\Theta_t^1\big) \, \mathrm{d}t + \beta \sum_{i=1}^k \partial_i \Theta\big(f_1(t, Y_t), \dots, f_k(t, Y_t)\big) f_{i+1}(t, Y_t) \, \mathrm{d}B_t.$$

Here the second-order term of Itô's formula produces Θ_t^0, and the drifts in (5.8) produce $\alpha(t)\Theta_t^1$. By Lemma 5.1, the integrand of the stochastic integral is bounded. Taking expectations, we obtain (5.11). The integrand in (5.11) is bounded by the supremum of $|\Theta^0| + |\Theta^1|$ over the box $\prod_{i \le k+1}[-C_i, C_i]$, which proves the Lipschitz bound.

Step 4: four moments. For $\Theta = x_1^2$, x_2^2, x_3^2, and x_2^3, the definition (5.7) gives

$$\begin{aligned}
(x_1^2)^0 &= bx_2^2, & (x_1^2)^1 &= 0,\\
(x_2^2)^0 &= bx_3^2, & (x_2^2)^1 &= -2bx_2^3,\\
(x_3^2)^0 &= bx_4^2, & (x_3^2)^1 &= -6bx_2x_3^2,\\
(x_2^3)^0 &= 3bx_2x_3^2, & (x_2^3)^1 &= -3bx_2^4.
\end{aligned}$$

For example, for $\Theta = x_3^2$ we have $\Theta^0 = (b/2) \cdot 2x_4^2$ and $\Theta^1 = -b \cdot 2x_3\big(\binom{2}{1}x_2x_3 + \binom{2}{2}x_3x_2\big)$. Inserting these into (5.11), we obtain (5.12). The integrand $\mathbb{E}\chi_r^2$ in the first identity is continuous in r by (ii), and the first identity implies that Γ is continuously differentiable with $\Gamma' = b\mathbb{E}\chi^2$. In the second identity, the functions $r \mapsto \mathbb{E}w_r^2$ and $r \mapsto \mathbb{E}\chi_r^3$ are continuous by (ii), and α is right-continuous. The integrand is therefore right-continuous, and it is continuous at every point of continuity of α. The fundamental theorem of calculus now gives (5.13). □

5.3 The Parisi distribution function on $(0, q_\beta)$

From now on, $\alpha = \alpha_\beta$ is the Parisi distribution function, which is unique by Theorem 4.5, μ_β is the Parisi measure, Y is the associated scalar diffusion, and u_t, χ_t, w_t are as in Lemma 5.2. By Theorem 4.7, $\operatorname{supp}\mu_\beta = [0, q_\beta]$ with $0 < q_\beta < 1$, and μ_β has an atom at q_β. We set

$$K_\beta = \frac{\beta^5}{2}\big(C_4^2 + 12C_3^2 + 6\big), \tag{5.14}$$

with C_3 and C_4 as in (5.2).

Lemma 5.3 (The Parisi distribution function on the support)**.** *The following hold.*

(i) *For all $t \in (0, q_\beta)$,*

$$\beta^2\mathbb{E}\chi(t, Y_t)^2 = 1, \qquad \mathbb{E}w(t, Y_t)^2 = 2\alpha(t)\mathbb{E}\chi(t, Y_t)^3. \tag{5.15}$$

(ii) *For all $t \in (0, q_\beta)$, we have $\mathbb{E}\chi_t^3 \geq \beta^{-3}$ and*

$$\alpha(t) = \frac{\mathbb{E}w_t^2}{2\mathbb{E}\chi_t^3}. \tag{5.16}$$

(iii) *We have $\alpha(t) > 0$ for all $t \in (0, 1]$. For $0 < s \leq t < q_\beta$,*

$$0 \leq \alpha(t) - \alpha(s) = \mu_\beta((s, t]) \leq K_\beta(t - s).$$

In particular, μ_β has no atoms in $(0, q_\beta)$.

(iv) *The function α is infinitely differentiable on $(0, q_\beta)$. For every polynomial Θ, the function $t \mapsto \mathbb{E}\Theta_t$ of Lemma 5.2 is infinitely differentiable on $(0, q_\beta)$. In particular, $t \mapsto \mathbb{E}\chi_t^3$ and $t \mapsto \alpha(t)^2\beta^6\mathbb{E}\chi_t^3$ are infinitely differentiable on $(0, q_\beta)$.*

(v) *The function Φ_α is infinitely differentiable on $(0, q_\beta) \times \mathbb{R}$. For every compact interval $I \subset (0, q_\beta)$ and all integers $i, j \geq 0$ with $i + j \geq 1$, there exists $C(\beta, I, i, j) < \infty$ such that*

$$\sup_{t \in I}|\alpha^{(i)}(t)| + \sup_{(t,z) \in I \times \mathbb{R}}|\partial_t^i\partial_z^j\Phi_\alpha(t, z)| \leq C(\beta, I, i, j).$$

The proof of (i) uses the identity $\Gamma(t) = t$, which is a property of α_β itself. The step distribution functions used in the proof of Lemma 5.2 do not satisfy it, and this is why we established the identities of Lemma 5.2 for every distribution function first.

Proof. We first obtain the two identities in (i): the first at every point, from $\Gamma(t) = t$, and the second at almost every point, from the constancy of $\mathbb{E}\chi_t^2$. Since α is right-continuous, the second identity then holds at every point, which gives the quotient formula (5.16). Using the quotient formula and the Lipschitz bounds of Lemma 5.2, we show that α is Lipschitz, and an induction on the order of differentiability gives its smoothness. The bounds on mixed derivatives then follow from the differentiated Parisi equation.

Step 1: the first identity. The first identity is also stated in Corollary 4.8; we recall the argument. By (4.10), $\Gamma(t) = t$ for $0 < t < q_\beta$. By Lemma 5.2 (iii), Γ is continuously differentiable with $\Gamma'(t) = b\mathbb{E}\chi_t^2$. Differentiating, we find that $b\mathbb{E}\chi_t^2 = \Gamma'(t) = 1$ for all $t \in (0, q_\beta)$, which is the first identity in (5.15).

Step 2: the second identity almost everywhere. By Step 1, $\mathbb{E}\chi_t^2 = \beta^{-2}$ is constant on $(0, q_\beta)$. The second identity in (5.12) implies that, for $0 < s < t < q_\beta$,

$$0 = \mathbb{E}\chi_t^2 - \mathbb{E}\chi_s^2 = b\int_s^t \bigl(\mathbb{E}w_r^2 - 2\alpha(r)\mathbb{E}\chi_r^3\bigr)\,\mathrm{d}r.$$

The integral of the bounded function $r \mapsto \mathbb{E}w_r^2 - 2\alpha(r)\mathbb{E}\chi_r^3$ vanishes over every subinterval of $(0, q_\beta)$. Lebesgue's differentiation theorem then implies that $\mathbb{E}w_r^2 = 2\alpha(r)\mathbb{E}\chi_r^3$ for almost every $r \in (0, q_\beta)$.

Step 3: the second identity at every point, and the quotient. Let $t \in (0, q_\beta)$. By Step 2, there exist $r_n \downarrow t$ such that $\mathbb{E}w_{r_n}^2 = 2\alpha(r_n)\mathbb{E}\chi_{r_n}^3$ for all n. Recall from Lemma 5.2 (ii) that the functions $r \mapsto \mathbb{E}w_r^2$ and $r \mapsto \mathbb{E}\chi_r^3$ are continuous, and that α is right-continuous. Letting $n \to \infty$, we find that $\mathbb{E}w_t^2 = 2\alpha(t)\mathbb{E}\chi_t^3$. This proves (i). Since $\chi_t > 0$ and $x \mapsto x^{3/2}$ is convex on $[0, \infty)$, Jensen's inequality and Step 1 imply that

$$\mathbb{E}\chi_t^3 = \mathbb{E}\bigl[(\chi_t^2)^{3/2}\bigr] \geq \bigl(\mathbb{E}\chi_t^2\bigr)^{3/2} = \beta^{-3}.$$

Dividing the second identity in (5.15) by $2\mathbb{E}\chi_t^3$, we obtain (5.16), which proves (ii).

Step 4: positivity and the Lipschitz bound. Since $0 \in \operatorname{supp}\mu_\beta$, every interval $[0, t]$ with $t > 0$ has positive μ_β-mass, so $\alpha(t) > 0$ for $t \in (0, 1]$. By Lemma 5.1, we have $0 < \chi \leq 1$, $|w| \leq C_3$, and $|f_4| \leq C_4$. Inserting these bounds into the last two identities in (5.12), we find that, for all $s, t \in [0, 1]$,

$$|\mathbb{E}w_t^2 - \mathbb{E}w_s^2| \leq b\bigl(C_4^2 + 6C_3^2\bigr)|t - s|, \qquad |\mathbb{E}\chi_t^3 - \mathbb{E}\chi_s^3| \leq 3b\bigl(C_3^2 + 1\bigr)|t - s|. \tag{5.17}$$

Let $0 < s \le t < q_\beta$. By (5.16) at t and the identity $\mathbb{E}w_s^2 = 2\alpha(s)\mathbb{E}\chi_s^3$ at s, we have

$$\alpha(t) - \alpha(s) = \frac{\mathbb{E}w_t^2 - \mathbb{E}w_s^2}{2\mathbb{E}\chi_t^3} - \alpha(s)\,\frac{\mathbb{E}\chi_t^3 - \mathbb{E}\chi_s^3}{\mathbb{E}\chi_t^3}.$$

Using (5.17), $\mathbb{E}\chi_t^3 \ge \beta^{-3}$, and $0 \le \alpha(s) \le 1$, we conclude that

$$|\alpha(t) - \alpha(s)| \le \beta^3\Big(\frac{b}{2}\big(C_4^2 + 6C_3^2\big) + 3b\big(C_3^2 + 1\big)\Big)(t - s) = K_\beta(t - s).$$

Since α is nondecreasing and $\alpha(t) - \alpha(s) = \mu_\beta((s,t])$, this proves (iii); the continuity of α on $(0, q_\beta)$ excludes atoms there.

Step 5: smoothness of α and of the moments. Let $I_0 = (0, q_\beta)$. We prove by induction on $i \ge 0$ that $\alpha \in C^i(I_0)$ and that $t \mapsto \mathbb{E}\Theta_t$ belongs to $C^{i+1}(I_0)$ for every polynomial Θ. By Step 4, α is continuous on I_0. Then the integrand $r \mapsto \mathbb{E}\Theta_r^0 + \alpha(r)\mathbb{E}\Theta_r^1$ in (5.11) is continuous on I_0, and $\mathbb{E}\Theta_t$ is continuously differentiable on I_0 by the fundamental theorem of calculus, with

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Theta_t = \mathbb{E}\Theta_t^0 + \alpha(t)\mathbb{E}\Theta_t^1 \qquad (t \in I_0). \tag{5.18}$$

This proves the claim for $i = 0$. Suppose that it holds for some $i \ge 0$. Since the moments $\mathbb{E}w_t^2$ and $\mathbb{E}\chi_t^3$ belong to $C^{i+1}(I_0)$ and $\mathbb{E}\chi_t^3 \ge \beta^{-3}$, the formula (5.16) implies that $\alpha \in C^{i+1}(I_0)$. By the induction hypothesis, the moments $\mathbb{E}\Theta_t^0$ and $\mathbb{E}\Theta_t^1$ belong to $C^{i+1}(I_0)$. Then the right side of (5.18) belongs to $C^{i+1}(I_0)$, and we obtain $\mathbb{E}\Theta_t \in C^{i+2}(I_0)$. This completes the induction and proves (iv).

Step 6: mixed derivatives. Let $j \ge 0$. By (5.4) and the continuity of α on I_0, the function $t \mapsto f_j(t,z)$ is continuously differentiable on I_0 for every z, and

$$\partial_t f_j = -\frac{b}{2}\big(f_{j+2} + \alpha\,\partial_z^j(u^2)\big) \qquad \text{on } I_0 \times \mathbb{R}. \tag{5.19}$$

The right side is a polynomial in α and $f_1, \dots, f_{j+2}$, and it is continuous on $I_0 \times \mathbb{R}$. We claim that, for every $i \ge 1$, the derivative $\partial_t^i f_j$ exists on $I_0 \times \mathbb{R}$ and is a polynomial, with coefficients depending only on b, i, and j, in $\alpha, \alpha', \dots, \alpha^{(i-1)}$ (evaluated at t) and $f_1, \dots, f_{j+2i}$ (evaluated at (t,z)). For $i = 1$, this is (5.19). If it holds for i, we differentiate the polynomial in t, using (iv) for the derivatives of α and (5.19) for the derivatives of $f_1, \dots, f_{j+2i}$, which gives the claim for $i+1$. Let $I \subset I_0$ be compact. On $I \times \mathbb{R}$, the functions $\alpha, \alpha', \dots, \alpha^{(i)}$ are bounded because they are continuous on I, and we have $|f_l| \le C_l$ by Lemma 5.1. We obtain the bound in (v), with a constant depending on β, I, i, and j. Finally, the right side of (5.19) for $j+1$ is the z-derivative of the right side for j, because $\partial_z \partial_z^j(u^2) = \partial_z^{j+1}(u^2)$. This shows that $\partial_z \partial_t f_j = \partial_t f_{j+1}$, and by Schwarz's

theorem and induction on i, we have $\partial_z \partial_t^i f_j = \partial_t^i f_{j+1}$. Every partial derivative of Φ_α, taken in any order, is therefore one of the continuous functions $\partial_t^i \partial_z^j \Phi_\alpha$, and Φ_α is infinitely differentiable on $I_0 \times \mathbb{R}$. This completes the proof. □

Remark 5.4. The second identity in (5.15) does not extend to $t = q_\beta$. Indeed, $t \mapsto \mathbb{E} w_t^2$ and $t \mapsto \mathbb{E}\chi_t^3$ are continuous on $[0,1]$, and letting $t \uparrow q_\beta$ in (5.15), we find that $\mathbb{E} w_{q_\beta}^2 = 2\alpha(q_\beta-)\mathbb{E}\chi_{q_\beta}^3$. However, $\mathbb{E}\chi_{q_\beta}^3 > 0$, and $\alpha(q_\beta-) = 1 - \mu_\beta(\{q_\beta\}) < 1 = \alpha(q_\beta)$ because μ_β has an atom at q_β. The barrier argument uses (5.15) and Lemma 5.3 only on compact subintervals of $(0, q_\beta)$.

We conclude with the positive constant that enters the cubic pressure deficit. The quantity bounded below in (5.20) is the scalar factor of the cubic term in the expansion of Proposition 7.13.

Corollary 5.5 (The constant m_β)**.** *Let $I_\beta \subset (0, q_\beta)$ be a nonempty compact interval, and set $m_\beta = \beta^3 \alpha(\min I_\beta)^2$. Then $m_\beta > 0$, and*

$$m_\beta \le \inf_{r \in I_\beta} \alpha(r)^2 \beta^6 \mathbb{E}\chi(r, Y_r)^3. \tag{5.20}$$

Proof. By Lemma 5.3 (iii), we have $\alpha(\min I_\beta) > 0$, so $m_\beta > 0$. Let $r \in I_\beta$. Since α is nondecreasing, we have $\alpha(r) \ge \alpha(\min I_\beta)$, and Lemma 5.3 (ii) gives $\mathbb{E}\chi(r, Y_r)^3 \ge \beta^{-3}$. Multiplying these bounds, we find that $\alpha(r)^2 \beta^6 \mathbb{E}\chi(r, Y_r)^3 \ge \beta^3 \alpha(\min I_\beta)^2 = m_\beta$. □

The constant m_β depends only on β and I_β. In the proof of Theorem 1.1, the interval I_β is fixed in Section 10, before any of the small parameters of the local comparison is chosen.

SECTION 6

Short-time estimates for Gaussian hierarchies

In this section, we prove the analytic estimates on Gaussian hierarchies that the three-replica covariance perturbation of Section 7 requires. Near the window where the covariance path is perturbed, the one-site recursion of that comparison is a composition of short semilinear parabolic flows on $\mathbb{R}^3$, one for each of the fixed number of stages of the perturbed path. We must differentiate its value four times in the width parameter h of the window and twice in the parameter $k = \delta/h$, uniformly in the starting field. Several covariance matrices of the path are singular, and the stage of the terminal correction has a duration proportional to hk^2, which vanishes at $k = 0$. We state and prove the estimates for a general finite composition of such flows on $\mathbb{R}^\ell$.

The section has four main results. Lemma 6.2 identifies the limit of finite Gaussian hierarchies (Definition 4.14) along refinements with the solution of a flow equation, with an error controlled by the L^1 distance between the masses. Proposition 6.3 bounds all mixed derivatives of the flow in the parameters (h, k) and in the field, uniformly in the field, including the endpoint $h = 0$ and stages of zero duration, and Lemma 6.7 gives the corresponding bounds for the penalty obtained by weighting the squared covariance by the mass. Lemma 6.13 and Corollary 6.14 control the response of a Gaussian hierarchy to a small bounded change of its terminal function, with an error that does not depend on the number of levels. In Section 7, we use the first three results to obtain Lemma 7.4, and Corollary 6.14 is applied to the part of the hierarchy that precedes the window, in the proof of Proposition 7.13. From earlier sections, we use the integration step and the hierarchies of Section 4 (Definitions 4.1 and 4.14 and Lemma 4.2). The regularity and stability of the scalar Parisi equation (Proposition 4.3) enter only the proof of Corollary 6.14. The remaining tools are the Gaussian interpolation formula, Itô's formula, and Gronwall's inequality [107, Section A.13].

For the scalar Parisi equation, the approximation by step profiles and the bounds on spatial derivatives that are uniform in the profile are due to Auffinger and Chen [12, Propositions 1–2]. The existence, uniqueness, and spatial smoothness of weak solutions were proved by Jagannath and Tobasco [62, Theorem 1]. In our normalization, these facts and the Lipschitz dependence on the profile are stated in Proposition 4.3, which is [77, Lemma 3.1]. Bovier and Klimovsky derived the matrix form of the Parisi equation for vector spins [29, Section 6, Proposition 6.1 and (6.5)]. The arguments below use the Feynman–Kac representation for the linear equations obtained by differentiating the flow.

We give the estimates for a vector equation whose diffusion matrices may be singular and whose coefficients depend on parameters, and we give in detail the bookkeeping of mixed derivatives in the parameters, which the later sections need.

Notation. Throughout this section, $\ell \geq 1$ is an integer ($\ell = 3$ in Section 7, one coordinate for each replica), $b > 0$ ($b = \beta^2$ in all applications), and $x \in \mathbb{R}^\ell$ denotes the field variable, with coordinates indexed by $i, i' \in \{1, \dots, \ell\}$. We write $\langle \cdot, \cdot \rangle$ and $|\cdot|$ for the Euclidean inner product and norm, $|\Sigma|$ for the operator norm of a matrix, $U : V = \sum_{i,i'} U_{ii'} V_{ii'}$ and $\|U\|_F^2 = U : U$, and $\Sigma \succeq 0$ when Σ is symmetric and positive semidefinite; $\Sigma^{1/2}$ is its positive semidefinite square root, and $\mathbf{g}$ is a standard Gaussian vector in $\mathbb{R}^\ell$. For a multi-index $\iota \in \mathbb{N}^\ell$, D^ι is the corresponding derivative in x and $|\iota|$ its order; ∇ and D^2 are the gradient and the Hessian in x. For $\varphi \colon \mathbb{R}^\ell \to \mathbb{R}$ and an integer $r \geq 1$, we set $\|D^r \varphi\|_\infty = \max_{|\iota|=r} \sup_x |D^\iota \varphi(x)|$, and for $2 \leq r \leq \infty$ we let $\mathcal{C}^r$ be the set of $\varphi \in C^r(\mathbb{R}^\ell)$ with $\|D^{r'} \varphi\|_\infty < \infty$ for all finite $1 \leq r' \leq r$. A function in $\mathcal{C}^r$ grows at most linearly and need not itself be bounded. Parameter derivatives are taken in two real variables (h, k): for $\gamma = (\gamma_h, \gamma_k) \in \mathbb{N}^2$, we write $\partial^\gamma = \partial_h^{\gamma_h} \partial_k^{\gamma_k}$, $|\gamma| = \gamma_h + \gamma_k$, $\gamma' \leq \gamma$ for the componentwise order, $\binom{\gamma}{\gamma'} = \binom{\gamma_h}{\gamma_h'} \binom{\gamma_k}{\gamma_k'}$, and $\gamma! = \gamma_h! \gamma_k!$. We say that the mixed derivative $\partial^\gamma D^\iota$ has *weight* $|\iota| + 2|\gamma|$. One parameter derivative counts as two spatial derivatives, because a parameter derivative of a covariance coefficient multiplies a second derivative in x.

6.1 Flows and the main estimates

Definition 6.1 (Stage data and flows)**.** Let $N \geq 1$. A *stage datum* consists of, for each $j \in \{1, \dots, N\}$, a matrix $B_j \succeq 0$, a duration $\lambda_j \geq 0$, and a measurable mass function $m_j \colon [0,1] \to [0,1]$. Stage j occupies the interval $[j-1, j]$ of the variable $\tau \in [0, N]$. For a terminal function $f \in \mathcal{C}^2$, the *flow equation* is

$$-\partial_\tau F = \frac{b\lambda_j}{2} \Big\{ B_j : D^2 F + m_j(\tau - j + 1) \big\langle \nabla F, B_j \nabla F \big\rangle \Big\} \quad \text{on } (j-1, j) \times \mathbb{R}^\ell, \tag{6.1}$$

for $j = 1, \dots, N$, with $F(N, \cdot) = f$ and F continuous on $[0, N] \times \mathbb{R}^\ell$. We call $F(0, \cdot)$ the *output* of the flow.

We solve the equation backward, from $\tau = N$ to $\tau = 0$, as the Parisi equation (4.4) is solved backward from time one. Stage 1 is the outermost stage, nearest to the root of the hierarchy, and stage N is the innermost. On stage j the Gaussian field has covariance $b\lambda_j B_j$ per unit of τ; in the notation of Definition 4.14, the overlap covariance path increases by $\lambda_j B_j$ over stage j. For example, the Parisi equation for a distribution function α on an interval $[t_1, t_2]$ is the case $\ell = 1$, $N = 1$, $B_1 = 1$, $\lambda_1 = t_2 - t_1$, and $m_1(\tau) = \alpha(t_1 + \tau(t_2 - t_1))$, with $b = \beta^2$.

Step flows. We first define solutions for masses that are step functions. A *step flow* is the flow of a stage datum in which, for each j, there exists a partition $0 = \tau_{j,0} < \tau_{j,1} < \cdots < \tau_{j,M_j} = 1$ such that $m_j = \mu_{j,i}$ on $[\tau_{j,i-1}, \tau_{j,i})$. With the integration step of Definition 4.1, we set $F(N, \cdot) = f$ and

$$F(\tau, \cdot) = T_{\mu_{j,i},\,(j-1+\tau_{j,i}-\tau)b\lambda_j B_j} F(j - 1 + \tau_{j,i}, \cdot) \quad \text{for } j - 1 + \tau_{j,i-1} \le \tau \le j - 1 + \tau_{j,i}, \tag{6.2}$$

proceeding from the piece (N, M_N) to the piece $(1, 1)$. The formulas for two consecutive pieces agree at their common endpoint, because $T_{m,0}\varphi = \varphi$. The output $F(0, x)$ is the value at x of a finite Gaussian hierarchy: its levels, listed from the root, are the pieces (j, i) in lexicographic order, and the level (j, i) has covariance $b\lambda_j B_j(\tau_{j,i} - \tau_{j,i-1})$ and is integrated at mass $\mu_{j,i}$. By Lemma 6.9 below, F is continuous and solves (6.1) on each piece, where the mass is constant.

If the masses m_j are continuous and $\mathfrak{p} = (\tau_{j,i})$ is a family of partitions, the *sampled masses* are $m_j^{\mathfrak{p}} = m_j(\tau_{j,i-1})$ on $[\tau_{j,i-1}, \tau_{j,i})$, and the *mesh* $|\mathfrak{p}|$ is the largest length $\tau_{j,i} - \tau_{j,i-1}$. If each m_j is nondecreasing and $m_j(1) \le m_{j+1}(0)$ for $j < N$, then the sampled masses are nondecreasing from the root, as in an admissible path of Definition 4.14.

Lemma 6.2 (Step approximation)**.** *Fix a stage datum with matrices B_j and durations λ_j, and a terminal function $f \in \mathcal{C}^2$. Let $L_f = \sup_x |\nabla f(x)|$.*

(a) *Let F and $\widetilde{F}$ be the step flows with step masses m_j and $\widetilde{m}_j$. Then $\sup_{\tau,x} |\nabla F(\tau, x)| \le L_f$ and*

$$\sup_{\tau,x} |F(\tau, x) - \widetilde{F}(\tau, x)| \le \frac{bL_f^2}{2} \sum_{j=1}^{N} \lambda_j |B_j| \int_0^1 |m_j(\tau) - \widetilde{m}_j(\tau)| \, \mathrm{d}\tau. \tag{6.3}$$

If F' is the step flow with the masses m_j and a terminal function $f' \in \mathcal{C}^2$ such that $f - f'$ is bounded, then $\sup_{\tau,x} |F - F'| \le \sup_x |f - f'|$.

(b) *Let m_j be measurable. For every sequence of step masses $m_j^{(l)}$ with $\int_0^1 |m_j^{(l)} - m_j| \, \mathrm{d}\tau \to 0$ for each j, the step flows converge uniformly on $[0, N] \times \mathbb{R}^\ell$ to a limit F that does not depend on the sequence. We call F the* flow *of the stage datum with terminal function f. The three estimates of* (a) *hold for these flows.*

(c) *Let $1 \le N' < N$. Let F_{in} be the flow of stages $N' + 1, \dots, N$ with terminal function f, and suppose that $f_{\mathrm{in}} = F_{\mathrm{in}}(N', \cdot)$ belongs to $\mathcal{C}^2$. Then F coincides on $[0, N'] \times \mathbb{R}^\ell$ with the flow of stages $1, \dots, N'$ with terminal function f_{in}.*

(d) *Suppose that $f \in \mathcal{C}^\infty$ and $2\ell b \|D^2 f\|_\infty \sum_j \lambda_j |B_j| \le 1$. Then the flow F of* (b) *has continuous and bounded spatial derivatives of every order $r \ge 1$, with $\|D^2 F(\tau, \cdot)\|_\infty \le 2\|D^2 f\|_\infty$, and, for the step flows $F^{(l)}$ of* (b) *and every $\iota \ne 0$,*

$$\sup_{\tau,x} |D^\iota F^{(l)}(\tau, x) - D^\iota F(\tau, x)| \le C \sum_{j=1}^{N} \lambda_j \int_0^1 |m_j^{(l)} - m_j| \, \mathrm{d}\tau, \tag{6.4}$$

where C depends only on ℓ, b, $|\iota|$, $\max_j|B_j|$, $\sum_j \lambda_j|B_j|$, and $\|D^r f\|_\infty$ for $1 \le r \le |\iota|+1$. If moreover each m_j is continuous on $[0,1]$, then F is continuously differentiable in τ on each closed stage $[j-1,j]$, with one-sided derivatives at its endpoints, and solves (6.1)*. It is the unique such solution with bounded first and second spatial derivatives that is of class $C^{1,2}$ on each closed stage.*

In words, the hierarchy values converge along refinements of the masses, at a rate given by the L^1 distance between the masses, and the limit is the solution of the flow equation; for short flows the same holds for every spatial derivative. Part (c) says that the limiting flows compose stage by stage. Parts (a)–(c) hold for stages of any duration. The constant in (6.3) involves only the gradient of f. To identify the gradient of the limit in (b), we also use the bound $K_f + L_f^2$ of Lemma 6.11(a) on the Hessians of the step flows, where $K_f = \sup_x |D^2 f(x)|$. This bound does not require short stages either. Part (d) is stated only for short flows, which are the only ones we need.

Parameter-dependent data. We now let the data depend on two parameters. Fix $N \ge 1$, $h_1 > 0$, and $k_0 > 0$, and let $\mathcal{U} = [0,h_1] \times [-k_0,k_0]$. A function on $\mathcal{U}$ is called *smooth* if all its partial derivatives exist and are continuous on $\mathcal{U}$; at boundary points, the derivatives are one-sided. In particular, derivatives in h at $h = 0$ are right derivatives. We assume the following.

(H1) For each j, the map $(h,k) \mapsto B_j(h,k)$ is smooth on $\mathcal{U}$, with values in the positive semidefinite matrices.

(H2) For each j, $\lambda_j(h,k) = h\kappa_j(h,k)$, where $\kappa_j \colon \mathcal{U} \to [0,\infty)$ is smooth. The function κ_j may vanish.

(H3) For each j, $m_j \colon [0,1] \times \mathcal{U} \to [0,1]$ is continuous, and all its derivatives in (h,k) exist and are continuous on $[0,1] \times \mathcal{U}$.

(H4) The terminal function $f \colon \mathcal{U} \times \mathbb{R}^\ell \to \mathbb{R}$ is smooth in the same sense, and $\partial^\gamma D^\iota f$ is bounded on $\mathcal{U} \times \mathbb{R}^\ell$ for every $(\gamma,\iota) \ne (0,0)$.

(H5) The number $h_0 \in (0,h_1]$ satisfies $2\ell b\, h_0 \Lambda_* \|D^2 f\|_\infty \le 1$, where $\Lambda_* = \max_{\mathcal{U}} \sum_j \kappa_j |B_j|$ and $\|D^2 f\|_\infty = \sup_{\mathcal{U}\times\mathbb{R}^\ell} \max_{i,i'} |\partial_{ii'} f|$.

Let $\mathcal{U}_0 = [0,h_0] \times [-k_0,k_0]$. For $(h,k) \in \mathcal{U}_0$, let $F(\tau,h,k,x)$ be the flow, in the sense of Lemma 6.2, of the stage datum $(B_j(h,k), \lambda_j(h,k), m_j(\cdot,h,k))_{j \le N}$ with terminal function $f(h,k,\cdot)$, and let $\mathcal{B}(h,k,x) = F(0,h,k,x)$ be its output.

Proposition 6.3 (Uniform bounds for short flows)**.** *Assume* (H1)–(H5)*.*

(i) *For each $(h,k) \in \mathcal{U}_0$, the function $F(\cdot,h,k,\cdot)$ is continuously differentiable in τ on each closed stage, infinitely differentiable in x, and it is the unique solution of* (6.1) *with bounded first and second spatial derivatives that is of class $C^{1,2}$ on each closed stage.*

(ii) *All partial derivatives $\partial^\gamma D^\iota F$ exist and are continuous on $[0,N]\times\mathcal{U}_0\times\mathbb{R}^\ell$. For every $(\gamma,\iota)\neq(0,0)$, there exists C such that*

$$\sup_{[0,N]\times\mathcal{U}_0\times\mathbb{R}^\ell}|\partial^\gamma D^\iota F|\le C,$$

where C depends only on ℓ, N, b, h_1, γ, and $|\iota|$, on the suprema over $\mathcal{U}$ of $|\partial^{\gamma'}(\kappa_j B_j)|$ and over $[0,1]\times\mathcal{U}$ of $|\partial^{\gamma'}(\kappa_j m_j B_j)|$, for $\gamma'\le\gamma$, and on the suprema over $\mathcal{U}\times\mathbb{R}^\ell$ of $|\partial^{\gamma'}D^{\iota'}f|$ for $\gamma'\le\gamma$ and $0<|\iota'|+2|\gamma'|\le|\iota|+2|\gamma|$. Moreover, $|F(\tau,h,k,x)|\le C_{\mathrm{grow}}(1+|x|)$, where C_{grow} depends only on ℓ, N, b, h_1, the supremum over $\mathcal{U}$ of $|\kappa_j B_j|$, the suprema over $\mathcal{U}\times\mathbb{R}^\ell$ of $|\nabla f|$ and $|D^2 f|$, and $\sup_{\mathcal{U}}|f(h,k,0)|$.

(iii) *For $\gamma\neq 0$, the derivative $\partial^\gamma F$ is continuously differentiable in τ on each closed stage, and it solves the equation obtained by applying ∂^γ to* (6.1)*, with terminal value $\partial^\gamma f$.*

(iv) *Let $\mathfrak{p}_l$ be partitions with $|\mathfrak{p}_l|\to 0$, and let $F^{\mathfrak{p}_l}$ be the step flows with the sampled masses $m_j^{\mathfrak{p}_l}(\cdot,h,k)$. Then $F^{\mathfrak{p}_l}\to F$ uniformly on $[0,N]\times\mathcal{U}_0\times\mathbb{R}^\ell$, and $\partial^\gamma D^\iota F^{\mathfrak{p}_l}\to\partial^\gamma D^\iota F$ uniformly on compact subsets of $[0,N]\times\mathcal{U}_0\times\mathbb{R}^\ell$, for every (γ,ι).*

To bound a derivative $\partial^\gamma D^\iota\mathcal{B}$, we use only derivatives of the terminal function of weight at most $|\iota|+2|\gamma|$ and (h,k)-derivatives of order at most γ of the coefficients. In Section 7, we need the derivatives $\partial_h^{\gamma_h}\partial_k^{\gamma_k}\mathcal{B}$ with $\gamma_h\le 4$ and $\gamma_k\le 2$ together with a few spatial derivatives; with at most eight spatial derivatives they have weight at most 20, and their bounds use (h,k)-derivatives of the masses and of the terminal function of order at most six. The value F itself grows at most linearly in x and is not estimated, except through its derivatives and through bounded differences such as $F(\tau,h,k,x)-F(\tau,h',k',x)$.

Remark 6.4 (The coefficients at $h=0$). Since $\lambda_j(0,k)=0$, we have $F(\tau,0,k,\cdot)=f(0,k,\cdot)$ for all τ, and hence $\partial_k^{\gamma_k}\mathcal{B}(0,k,\cdot)=\partial_k^{\gamma_k}f(0,k,\cdot)$ for every $\gamma_k\ge 0$. Writing $\bar m_j(k)=\int_0^1 m_j(\tau,0,k)\,\mathrm{d}\tau$ and evaluating f, B_j and κ_j at $(0,k)$, we have

$$\partial_h\mathcal{B}(0,k,x)=\partial_h f(0,k,x)+\frac{b}{2}\sum_{j=1}^N\kappa_j\Big\{B_j:D^2 f(0,k,x)+\bar m_j\big\langle\nabla f(0,k,x),B_j\nabla f(0,k,x)\big\rangle\Big\}. \tag{6.5}$$

Indeed, at $h=0$ the equation of Proposition 6.3(iii) for $\gamma=(1,0)$ has no terms involving $\partial_h F$, since the coefficients of (6.1) vanish, and its source is the derivative in h of the right side of (6.1) at $F=f(0,k,\cdot)$. Integrating it over $\tau\in[0,N]$, we obtain (6.5).

Remark 6.5 (Singular matrices, zero durations, and regularization). The hypotheses (H1)–(H5) and the constants in Proposition 6.3 involve no lower bound on B_j or on κ_j. In the proof, every Gaussian integral is parametrized by its covariance matrix $t\,bh\kappa_j(h,k)B_j(h,k)$, which is a smooth function of (h,k) with values in the positive semidefinite matrices,

including where it vanishes, and Lemma 6.8 differentiates Gaussian integrals as functions of the covariance. No square root of a covariance or of a duration is differentiated. For instance, a stage with $\kappa_j(h,k) = ck^2$ has zero duration at $k = 0$, and its contribution is differentiable in k there through the polynomial hk^2. One may also replace each B_j by $B_j + \varepsilon\mathrm{Id}$ with $0 < \varepsilon \le 1$. Let $K_* = \max_{\mathcal{U}} \sum_j \kappa_j$, and decrease h_0, if necessary, so that $2\ell b\, h_0(\Lambda_* + K_*)\|D^2 f\|_\infty \le 1$. The regularized data then satisfy (H1)–(H5), with upper bounds uniform in ε. For parameters with $h\kappa_j(h,k) > 0$, the diffusion matrix of stage j is positive definite. Stages with $h\kappa_j(h,k) = 0$, in particular all stages at $h = 0$, still have zero diffusion, and the regularized equations are not uniformly parabolic on $\mathcal{U}_0$. The regularized flows satisfy the bounds of Proposition 6.3 uniformly in ε, and they converge to F as $\varepsilon \to 0$, together with all derivatives $\partial^\gamma D^\iota$, uniformly on compact sets. Indeed, the difference between the regularized flow and F satisfies the linear equation in the proof of Lemma 6.2(a), with $B_j + \varepsilon\mathrm{Id}$ in place of B_j and with equal masses. Its source is $(b\lambda_j\varepsilon/2)\{\mathrm{Id} : D^2F + m_j|\nabla F|^2\}$, which is bounded by $C\varepsilon$. Then Lemma 6.10 shows that the values converge, and the compactness argument in the proof of Proposition 6.3(iv) identifies the limits of all derivatives. We do not use the regularization.

Remark 6.6 (Further parameters and restriction to a line). If the data depend on a further parameter in which no derivatives are needed (in Section 7, the left endpoint of the window), and the suprema listed in Proposition 6.3(ii) and the number $\Lambda_*\|D^2 f\|_\infty$ are bounded uniformly in it, then the conclusions hold uniformly in that parameter; for the linear growth bound, $\sup_{\mathcal{U}}|f(h,k,0)|$ must also be bounded uniformly in that parameter. The statements and the proofs of this section hold without change if k is a vector of parameters ranging over a product of compact intervals, with multi-indices in place of γ_k; a further parameter in which derivatives are needed can be included in this way. For a vector $e \in \mathbb{R}^\ell$ and $z \in \mathbb{R}$, the restriction $z \mapsto \mathcal{B}(h,k,ze)$ satisfies

$$\partial^\gamma \partial_z^r \big[\mathcal{B}(h,k,ze)\big] = \sum_{|\iota|=r} \frac{r!}{\iota!}\, e^\iota\, \partial^\gamma D^\iota \mathcal{B}(h,k,ze),$$

where $\iota! = \prod_i \iota_i!$ and $e^\iota = \prod_i e_i^{\iota_i}$. Its derivatives obey the bounds of Proposition 6.3(ii), multiplied by $(\sum_i |e_i|)^r$.

The last estimate of this section concerns the penalty. For a smooth function $Q_\star : \mathcal{U} \to \mathbb{R}^{\ell\times\ell}$ with symmetric values, let

$$Q(\tau) = Q_\star + \sum_{j'<j} \lambda_{j'} B_{j'} + (\tau - j + 1)\lambda_j B_j \qquad (j-1 \le \tau \le j),$$

where all functions are evaluated at (h,k). This is the overlap covariance path through the stages, started at $Q_\star$. We set

$$\Pi(h,k) = \sum_{j=1}^{N} \int_{j-1}^{j} m_j(\tau - j + 1)\, \partial_\tau \|Q(\tau)\|_F^2 \,\mathrm{d}\tau = 2 \sum_{j=1}^{N} \lambda_j \int_0^1 m_j(\tau)\, Q(j-1+\tau) : B_j \,\mathrm{d}\tau. \tag{6.6}$$

This is the contribution of the stages to the penalty $\int m \,\mathrm{d}\|Q\|_F^2$ of (4.14). For partitions $\mathfrak{p}$, the corresponding contribution of the step hierarchy with the sampled masses is

$$\Pi^{\mathfrak{p}}(h,k) = \sum_{j=1}^{N} \sum_{i=1}^{M_j} m_j(\tau_{j,i-1}) \Big(\|Q(j-1+\tau_{j,i})\|_F^2 - \|Q(j-1+\tau_{j,i-1})\|_F^2 \Big).$$

Lemma 6.7 (The penalty)**.** *Assume* (H1)–(H3). *Then* Π *is smooth on* $\mathcal{U}$*, and for every* γ *we have* $\sup_{\mathcal{U}} |\partial^\gamma \Pi| \le C$*, where* C *depends only on* ℓ*,* N*,* h_1*,* γ*, and the suprema over* $\gamma' \le \gamma$ *of* $|\partial^{\gamma'} Q_\star|$*,* $|\partial^{\gamma'}(\kappa_j B_j)|$ *and* $|\partial^{\gamma'}(\kappa_j m_j B_j)|$*. Moreover* $\sup_{\mathcal{U}} |\partial^\gamma(\Pi^{\mathfrak{p}} - \Pi)| \to 0$ *as* $|\mathfrak{p}| \to 0$*, for every* γ*.*

Proof. Since $m_j^{\mathfrak{p}}$ is constant on each piece, we have $\Pi^{\mathfrak{p}} = 2\sum_j \lambda_j \int_0^1 m_j^{\mathfrak{p}}(\tau)\, Q(j-1+\tau) : B_j \,\mathrm{d}\tau$, which is (6.6) with m_j replaced by $m_j^{\mathfrak{p}}$. The integrand on stage j is $Q(j-1+\tau) : (h\kappa_j m_j B_j)$, where $Q(j-1+\tau)$ is a polynomial in τ, $Q_\star$, and the matrices $h\kappa_{j'} B_{j'}$, and the (h,k)-derivatives of $m_j^{\mathfrak{p}}$ are those of m_j evaluated at the left endpoints of the pieces. Differentiating under the integral sign, we find that Π and $\Pi^{\mathfrak{p}}$ are smooth with the stated bounds. By (H3), every (h,k)-derivative of m_j is uniformly continuous on the compact set $[0,1] \times \mathcal{U}$, and we have $\partial^{\gamma'} m_j^{\mathfrak{p}} \to \partial^{\gamma'} m_j$ uniformly as $|\mathfrak{p}| \to 0$. This implies the last statement. □

6.2 Three tools

With the first tool, we differentiate a Gaussian integral with respect to its covariance, including at singular covariances.

Lemma 6.8 (Gaussian integrals as functions of the covariance)**.** *Let* Ω *be a convex subset of a Euclidean space. Let* $\mathcal{F} \colon \Omega \times \mathbb{R}^\ell \to \mathbb{R}$ *be continuous and twice continuously differentiable in its second variable* y*, with* $D_y^2 \mathcal{F}$ *continuous on* $\Omega \times \mathbb{R}^\ell$*, and suppose that there exists* $c < \infty$ *such that*

$$|\mathcal{F}(\omega,y)| + |\nabla_y \mathcal{F}(\omega,y)| + |D_y^2 \mathcal{F}(\omega,y)| \le c\, e^{c|y|} \qquad (\omega \in \Omega,\ y \in \mathbb{R}^\ell). \tag{6.7}$$

(a) *The map* $(\omega, \Sigma) \mapsto \mathbb{E}\mathcal{F}(\omega, \Sigma^{1/2}\mathbf{g})$ *is continuous on* $\Omega \times \{\Sigma \succeq 0\}$*. This holds for every continuous* $\mathcal{F}$ *with* $|\mathcal{F}| \le c e^{c|y|}$*.*

(b) *For $\omega \in \Omega$ and $\Sigma, \Sigma' \succeq 0$, with $\Sigma_t = \Sigma + t(\Sigma' - \Sigma)$,*

$$\mathbb{E}\mathcal{F}(\omega, \Sigma'^{1/2}\mathbf{g}) - \mathbb{E}\mathcal{F}(\omega, \Sigma^{1/2}\mathbf{g}) = \frac{1}{2}\int_0^1 (\Sigma' - \Sigma) : \mathbb{E}D_y^2\mathcal{F}(\omega, \Sigma_t^{1/2}\mathbf{g})\,\mathrm{d}t.$$

(c) *Let $\omega \mapsto \Sigma(\omega) \succeq 0$ be continuously differentiable on Ω, and suppose that $\partial_{\omega_a}\mathcal{F}$ exists, is continuous, and satisfies $|\partial_{\omega_a}\mathcal{F}| \le ce^{c|y|}$. Then the function $\omega \mapsto \mathbb{E}\mathcal{F}(\omega, \Sigma(\omega)^{1/2}\mathbf{g})$ has a continuous partial derivative in ω_a, one-sided at boundary points of Ω, and*

$$\partial_{\omega_a}\mathbb{E}\mathcal{F}(\omega, \Sigma(\omega)^{1/2}\mathbf{g}) = \mathbb{E}\partial_{\omega_a}\mathcal{F}(\omega, \Sigma(\omega)^{1/2}\mathbf{g}) + \frac{1}{2}\,\partial_{\omega_a}\Sigma(\omega) : \mathbb{E}D_y^2\mathcal{F}(\omega, \Sigma(\omega)^{1/2}\mathbf{g}). \tag{6.8}$$

Proof. For (a), recall that the positive square root is continuous on the positive semidefinite matrices. Indeed, if the spectra of Σ and Σ' lie in $[0, c']$, then $|\Sigma^{1/2} - \Sigma'^{1/2}| \le 2\sup_{[0,c']}|\sqrt{s} - P(s)| + |P(\Sigma) - P(\Sigma')|$ for every polynomial P, and $\sqrt{s}$ is a uniform limit of polynomials on $[0, c']$. If $(\omega_n, \Sigma_n) \to (\omega, \Sigma)$, then $\mathcal{F}(\omega_n, \Sigma_n^{1/2}\mathbf{g}) \to \mathcal{F}(\omega, \Sigma^{1/2}\mathbf{g})$ almost surely, and these random variables are dominated by $c\exp(cc'^{1/2}|\mathbf{g}|)$, where c' bounds the spectra of all Σ_n. By dominated convergence, the expectations converge, which proves (a).

For (b), let $\mathbf{g}'$ be a standard Gaussian vector independent of $\mathbf{g}$, and let $\zeta(t) = \sqrt{t}\,\Sigma'^{1/2}\mathbf{g}' + \sqrt{1-t}\,\Sigma^{1/2}\mathbf{g}$. The vector $\zeta(t)$ is centered Gaussian with covariance Σ_t, so it has the law of $\Sigma_t^{1/2}\mathbf{g}$. By the Gaussian interpolation formula [107, Lemma 1.3.1], for $0 < t < 1$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\mathcal{F}(\omega, \zeta(t)) = \frac{1}{2}(\Sigma' - \Sigma) : \mathbb{E}D_y^2\mathcal{F}(\omega, \zeta(t)).$$

That lemma is stated for infinitely differentiable functions whose derivatives grow more slowly than $e^{a|y|^2}$ for every $a > 0$. For $\mathcal{F}(\omega, \cdot)$ as in (6.7), we apply it to the convolution of $\mathcal{F}(\omega, \cdot)$ with a smooth compactly supported mollifier, integrate the identity between two times in $(0, 1)$, and let the mollifier shrink; both sides converge by dominated convergence. The right side of the displayed identity is continuous on $[0, 1]$ by (a), applied to $D_y^2\mathcal{F}$. Integrating over $t \in [0, 1]$, we obtain (b).

For (c), fix ω and write $\Sigma = \Sigma(\omega)$ and $\Sigma_\eta = \Sigma(\omega + \eta e_a)$ for $\omega + \eta e_a \in \Omega$, where e_a is the ath coordinate vector. Then

$$\begin{aligned}\mathbb{E}\mathcal{F}(\omega + \eta e_a, \Sigma_\eta^{1/2}\mathbf{g}) - \mathbb{E}\mathcal{F}(\omega, \Sigma^{1/2}\mathbf{g}) &= \int_0^\eta \mathbb{E}\partial_{\omega_a}\mathcal{F}(\omega + \eta' e_a, \Sigma_{\eta'}^{1/2}\mathbf{g})\,\mathrm{d}\eta' \\ &\quad + \frac{1}{2}\int_0^1 (\Sigma_\eta - \Sigma) : \mathbb{E}D_y^2\mathcal{F}\big(\omega, (\Sigma + t(\Sigma_\eta - \Sigma))^{1/2}\mathbf{g}\big)\,\mathrm{d}t,\end{aligned}$$

where we used the fundamental theorem of calculus and Fubini's theorem for the first term and (b) for the second; the segment from Σ to Σ_η consists of positive semidefinite matrices. We divide by η and let $\eta \to 0$. By (a), applied to $\partial_{\omega_a}\mathcal{F}$ and to the entries of

$D_y^2\mathcal{F}$, the two integrands converge uniformly to their values at $\eta = 0$, and we obtain (6.8). The right side of (6.8) is continuous by (a). □

The second tool collects the properties of one integration step $T_{m,\Sigma}$ of Definition 4.1 that we need. Recall that monotonicity, the translation identity, and the semigroup property are Lemma 4.2(a) and (b). For constant mass, the exponential change of variables that turns the flow into the heat equation is the Cole–Hopf transformation; its scalar Parisi form is recorded in [12, Section 2, (10)–(12)]. The next lemma gives the corresponding finite-dimensional Gaussian step.

Lemma 6.9 (One integration step)**.** *Let $\Sigma \succeq 0$, let $m \in [0,1]$, and let $\varphi \in \mathcal{C}^2$.*

(a) *We have $T_{m,\Sigma}\varphi \in \mathcal{C}^2$ and $\sup_x|\nabla T_{m,\Sigma}\varphi(x)| \le \sup_x|\nabla\varphi(x)|$. If $\varphi \in \mathcal{C}^r$, then $T_{m,\Sigma}\varphi \in \mathcal{C}^r$.*

(b) *The function $\psi(t,x) = T_{m,t\Sigma}\varphi(x)$, $t \ge 0$, is continuous and continuously differentiable in t, with a right derivative at $t = 0$; its first and second derivatives in x are continuous in (t,x); and*

$$\partial_t\psi = \frac{1}{2}\Big\{\Sigma : D^2\psi + m\big\langle\nabla\psi, \Sigma\nabla\psi\big\rangle\Big\}. \tag{6.9}$$

(c) *Let Ω be a compact convex subset of a Euclidean space, and let $\Sigma(\omega) \succeq 0$, $m(\omega) \in [0,1]$, and $\varphi(\omega,x)$ be smooth functions of $\omega \in \Omega$ and of $(\omega,x) \in \Omega \times \mathbb{R}^\ell$, such that $\partial_\omega^\gamma D^\iota\varphi$ is bounded for every $(\gamma,\iota) \ne (0,0)$. Then $\psi(t,\omega,x) = T_{m(\omega),t\Sigma(\omega)}\varphi(\omega,\cdot)(x)$ is infinitely differentiable in $(t,\omega,x) \in [0,1] \times \Omega \times \mathbb{R}^\ell$ (one-sided at the boundary), the function $\psi - \varphi$ is bounded, and every derivative $\partial_\omega^\gamma D^\iota\psi$ with $(\gamma,\iota) \ne (0,0)$ is bounded on $[0,1] \times \Omega \times \mathbb{R}^\ell$.*

Proof. Let $L_\varphi = \sup_x|\nabla\varphi(x)|$, and, for fixed x, let $\mathcal{Y}_x = \varphi(x + \Sigma^{1/2}\mathbf{g}) - \varphi(x)$, so that $|\mathcal{Y}_x| \le L_\varphi|\Sigma^{1/2}\mathbf{g}|$. For $m > 0$, we have $T_{m,\Sigma}\varphi(x) = \varphi(x) + m^{-1}\log\mathbb{E}e^{m\mathcal{Y}_x}$.

For (a), since the derivatives of $x \mapsto e^{m\varphi(x+y)-m\varphi(x)}$ of order at most two are bounded by $C(1+|y|)^2e^{L_\varphi|y|}$, we may differentiate under the expectation. We obtain

$$\nabla T_{m,\Sigma}\varphi(x) = \frac{\mathbb{E}\big[\nabla\varphi(x + \Sigma^{1/2}\mathbf{g})\, e^{m\mathcal{Y}_x}\big]}{\mathbb{E}e^{m\mathcal{Y}_x}}. \tag{6.10}$$

The right side is an average of values of $\nabla\varphi$ with respect to a probability measure, so its norm is at most L_φ. Differentiating further, we find that $D^\iota T_{m,\Sigma}\varphi(x)$, for $1 \le |\iota| \le r$, is a polynomial, with coefficients depending only on ι and m, in the averages $\mathbb{E}[P(x + \Sigma^{1/2}\mathbf{g})e^{m\mathcal{Y}_x}]/\mathbb{E}e^{m\mathcal{Y}_x}$ of products P of the derivatives $D^{\iota'}\varphi$ with $1 \le |\iota'| \le |\iota|$. These derivatives are bounded, and so are the averages. The case $m = 0$ is the same without the exponential weights.

For (b) and $m > 0$, let $U(t,x) = \mathbb{E}\exp\{m\varphi(x + (t\Sigma)^{1/2}\mathbf{g})\}$. The function $y \mapsto e^{m\varphi(x+y)}$ and its first two derivatives are bounded by $Ce^{m|\varphi(x)|+L_\varphi|y|}$, locally uniformly in x. By Lemma 6.8(a) and (b), with the covariances $t\Sigma$ and $t'\Sigma$, the function U is continuous, continuously differentiable in t, and $\partial_t U = (1/2)\Sigma : D^2U$. By Lemma 6.8(a), ∇U and D^2U are continuous in (t,x). Writing $\psi = m^{-1}\log U$, we have $\nabla U = mU\nabla\psi$ and

$D^2U = mU\{D^2\psi + m\nabla\psi\nabla\psi^{\mathsf{T}}\}$, and (6.9) follows. For $m = 0$, we apply Lemma 6.8(b) to φ directly.

For (c), we write $m = m(\omega)$, $\Sigma = \Sigma(\omega)$, and $\mathcal{Y} = \varphi(\omega, x + (t\Sigma)^{1/2}\mathbf{g}) - \varphi(\omega, x)$. For every $m \geq 0$,

$$\psi(t,\omega,x) - \varphi(\omega,x) = \int_0^1 \frac{\mathbb{E}\big[\mathcal{Y}\, e^{\rho m \mathcal{Y}}\big]}{\mathbb{E} e^{\rho m \mathcal{Y}}}\, \mathrm{d}\rho. \tag{6.11}$$

Indeed, for $m > 0$ the integrand is the derivative of $\rho \mapsto m^{-1} \log \mathbb{E} e^{\rho m \mathcal{Y}}$, and for $m = 0$ both sides equal $\mathbb{E}\mathcal{Y}$. We write $\widehat{\varphi}(\omega, x, y) = \varphi(\omega, x + y) - \varphi(\omega, x)$. The denominator is the Gaussian integral $\mathbb{E}\mathcal{F}(\omega, x, \rho; (t\Sigma)^{1/2}\mathbf{g})$ of the function

$$\mathcal{F}(\omega, x, \rho; y) = \exp\big\{\rho\, m(\omega)\, \widehat{\varphi}(\omega, x, y)\big\},$$

and the numerator is the Gaussian integral of $\widehat{\varphi}\,\mathcal{F}$. Every derivative of $\mathcal{F}$ and of $\widehat{\varphi}\,\mathcal{F}$ in (ω, x, ρ, y) is a polynomial in $\varphi(\omega, x + y) - \varphi(\omega, x)$, in the derivatives of m, and in the derivatives $\partial_\omega^{\gamma'} D^{\iota'} \varphi$ with $(\gamma', \iota') \neq (0, 0)$ evaluated at $(\omega, x + y)$ and (ω, x), times the exponential factor. Since $|\varphi(\omega, x + y) - \varphi(\omega, x)| \leq L_\varphi |y|$ with $L_\varphi = \sup|\nabla_x \varphi|$, all these derivatives are bounded by $C(1 + |y|)^C e^{L_\varphi |y|}$, uniformly in (t, ω, x, ρ). The covariance $(t, \omega) \mapsto t\Sigma(\omega)$ is smooth with positive semidefinite values. Applying Lemma 6.8(c) repeatedly, with the parameter (t, ω, x, ρ) in the convex set $[0, 1] \times \Omega \times \mathbb{R}^\ell \times [0, 1]$, we find that the numerator and the denominator are infinitely differentiable, and that each of their derivatives is a Gaussian integral of a function with the same bound; all these derivatives are bounded. The denominator is at least $\mathbb{E} e^{-L_\varphi |(t\Sigma)^{1/2}\mathbf{g}|}$, which is bounded below by a positive constant uniformly in $t \leq 1$ and $\omega \in \Omega$, since Ω is compact. Then the right side of (6.11) is infinitely differentiable, with all its derivatives and the function itself bounded. The derivatives of φ of order at least one are bounded by hypothesis, and this proves (c). □

The third tool is a bound for linear equations of the type (6.1). Its proof uses the stochastic representation of the solution and needs no lower bound on the diffusion matrix. If the diffusion matrix is degenerate, the representing diffusion moves in fewer directions.

Lemma 6.10 (A linear estimate)**.** *Let $T_* > 0$ and suppose that $[0, T_*]$ is the union of finitely many closed intervals with disjoint interiors. Let $\mathcal{A}(t) \succeq 0$ be constant on each of them. Let $\phi\colon [0, T_*] \times \mathbb{R}^\ell \to \mathbb{R}^\ell$ and $\mathrm{Y}\colon [0, T_*] \times \mathbb{R}^\ell \to \mathbb{R}$ be bounded and measurable, with ϕ Lipschitz in x uniformly in t, and with $t \mapsto \sup_x |\mathrm{Y}(t, x)|$ measurable. Let $\psi\colon [0, T_*] \times \mathbb{R}^\ell \to \mathbb{R}$ be continuous, with $|\psi(t, x)| + |\nabla\psi(t, x)| \leq c(1 + |x|)$, and suppose that on each of the intervals ψ coincides with a function of class $C^{1,2}$ and satisfies*

$$-\partial_t \psi = \mathcal{A}(t) : D^2\psi + \langle \phi, \nabla\psi \rangle + \mathrm{Y}$$

in its interior. Then, for every $t \in [0, T_*]$,

$$\sup_x |\psi(t,x)| \le \sup_x |\psi(T_*,x)| + \int_t^{T_*} \sup_x |\Upsilon(t',x)|\,\mathrm{d}t'. \tag{6.12}$$

Proof. Fix t and x, let W be a standard Brownian motion in $\mathbb{R}^\ell$, and let $\mathcal{X}$ solve

$$\mathcal{X}_{t'} = x + \int_t^{t'} \phi(t'', \mathcal{X}_{t''})\,\mathrm{d}t'' + \int_t^{t'} \bigl(2\mathcal{A}(t'')\bigr)^{1/2}\,\mathrm{d}W_{t''}, \qquad t \le t' \le T_*.$$

Since the noise term does not depend on $\mathcal{X}$ and ϕ is bounded and Lipschitz in space, this equation has a unique continuous adapted solution, obtained pathwise by Picard iteration and Gronwall's inequality. On each of the finitely many intervals, Itô's formula [64, Theorem 3.3.6] applies to the $C^{1,2}$ function that coincides with ψ there. Adding the results and using the equation for ψ, we obtain

$$\psi(T_*, \mathcal{X}_{T_*}) - \psi(t,x) = -\int_t^{T_*} \Upsilon(t', \mathcal{X}_{t'})\,\mathrm{d}t' + \int_t^{T_*} \bigl\langle \nabla\psi(t', \mathcal{X}_{t'}), \bigl(2\mathcal{A}(t')\bigr)^{1/2}\,\mathrm{d}W_{t'} \bigr\rangle.$$

We have $\mathbb{E}\sup_{t'} |\mathcal{X}_{t'}|^2 < \infty$ because ϕ is bounded, and the stochastic integral has mean zero because $\nabla\psi$ has linear growth. Taking expectations, we find

$$\psi(t,x) = \mathbb{E}\psi(T_*, \mathcal{X}_{T_*}) + \mathbb{E}\int_t^{T_*} \Upsilon(t', \mathcal{X}_{t'})\,\mathrm{d}t',$$

and (6.12) follows. This is the Feynman–Kac representation [64, Theorem 5.7.6], which we derived directly because ψ is of class $C^{1,2}$ only on each of the intervals. □

6.3 Spatial derivatives and the step limit

We first bound the spatial derivatives of step flows. The constants depend only on the durations and the matrices, not on the masses or the partitions.

Lemma 6.11 (Spatial bounds for step flows)**.** *Let F be a step flow with terminal function f.*

(a) *If $f \in \mathcal{C}^2$, then, for every τ,*

$$\sup_x |\nabla F(\tau,x)| \le L_f, \qquad \sup_x |D^2 F(\tau,x)| \le K_f + L_f^2,$$

where $L_f = \sup_x |\nabla f(x)|$, $K_f = \sup_x |D^2 f(x)|$, and $|\cdot|$ is the operator norm. This holds for stages of any duration and any masses in $[0,1]$.

(b) *If $f \in \mathcal{C}^\infty$ and $2\ell b \|D^2 f\|_\infty \sum_j \lambda_j |B_j| \le 1$, then $\|D^2 F(\tau,\cdot)\|_\infty \le 2\|D^2 f\|_\infty$ for every τ.*

(c) *Under the hypotheses of* (b)*, for every $r \ge 3$ there exists C, depending only on ℓ, b, r, $\sum_j \lambda_j |B_j|$, and $\|D^{r'} f\|_\infty$ for $2 \le r' \le r$, such that $\|D^r F(\tau,\cdot)\|_\infty \le C$ for every τ.*

Proof. For (a), we use the weighted martingale-variance decomposition in the proof of [77, Lemma 3.1], with a general vector terminal function in place of the scalar terminal function $\log(2\cosh z)$. The gradient bound in (a) follows from Lemma 6.9(a), applied to each step of (6.2). For the Hessian, fix (τ, x) and a unit vector v. By (6.2), $F(\tau, x)$ is the root value of a finite hierarchy, whose levels are the pieces after τ; if τ lies inside a piece, the first level is the remaining part of that piece. We list the levels from the root as $1, \dots, M$, with covariances Σ_j and masses $\mu_j \in [0, 1]$, and let $\Psi_M = f$ and $\Psi_{j-1} = T_{\mu_j, \Sigma_j} \Psi_j$, so that $\Psi_0 = F(\tau, \cdot)$. Each Ψ_j belongs to $\mathcal{C}^2$ by Lemma 6.9(a). Let $z_0 = x$, and given $z_0, \dots, z_{j-1}$, let z_j have the law $\mathcal{K}_{\mu_j, \Sigma_j, \Psi_j}(z_{j-1}, \cdot)$ of (4.2). The weights in (6.10) are the densities of this kernel, and therefore $\partial_v \Psi_{j-1}(z) = \mathbb{E}[\partial_v \Psi_j(z_j) \mid z_{j-1} = z]$. Differentiating (6.10) once more in the direction v, under the expectation as in the proof of Lemma 6.9(a), we find that

$$\partial_v^2 \Psi_{j-1}(z) = \mathbb{E}\big[\partial_v^2 \Psi_j(z_j) \;\big|\; z_{j-1} = z\big] + \mu_j \operatorname{Var}\big(\partial_v \Psi_j(z_j) \;\big|\; z_{j-1} = z\big).$$

The process $\mathcal{M}_j = \partial_v \Psi_j(z_j)$ is a bounded martingale, by the Markov property, and the conditional variance in the display, evaluated at $z = z_{j-1}$, has expectation $\mathbb{E}(\mathcal{M}_j - \mathcal{M}_{j-1})^2$. Iterating the display from $j = 1$ to $j = M$, we obtain

$$\partial_v^2 F(\tau, x) = \mathbb{E}\, \partial_v^2 f(z_M) + \sum_{j=1}^{M} \mu_j \mathbb{E}\big(\mathcal{M}_j - \mathcal{M}_{j-1}\big)^2.$$

Since $0 \le \mu_j \le 1$ and the increments of a martingale are orthogonal, the sum lies between 0 and $\sum_j \mathbb{E}(\mathcal{M}_j - \mathcal{M}_{j-1})^2 = \operatorname{Var}(\mathcal{M}_M) \le L_f^2$. We obtain $-K_f \le \partial_v^2 F(\tau, x) \le K_f + L_f^2$ for every unit vector v, and since $D^2 F(\tau, x)$ is symmetric, this is the Hessian bound. The bound does not depend on M, on the covariances, or on the order of the masses. This is the argument of the proof of Lemma 6.13 below, with a derivative in the field in place of a derivative in the terminal perturbation.

For (b) and (c), we note that, by Lemma 6.9(c) with a one-point parameter set, $F(\tau, \cdot) \in \mathcal{C}^\infty$ for every τ, all spatial derivatives of F are continuous, and on each closed piece they are continuously differentiable in τ and satisfy the equations obtained by differentiating (6.1) in x. On each piece, each spatial derivative of F is bounded. For $r \ge 1$, let $\Theta_r(\tau) = \max_{|\iota| = r} \sup_x |D^\iota F(\tau, x)|$. It is bounded on $[0, N]$ and lower semicontinuous, hence measurable. On stage j we write $\lambda(\tau)|B(\tau)| = \lambda_j |B_j|$.

Differentiating (6.1) in x_i and $x_{i'}$, we find that $\psi = \partial_{ii'} F$ satisfies, on stage j,

$$-\partial_\tau \psi = \frac{b\lambda_j}{2} B_j : D^2 \psi + b\lambda_j m_j \big\langle B_j \nabla F, \nabla \psi \big\rangle + b\lambda_j m_j \big\langle \nabla \partial_i F, B_j \nabla \partial_{i'} F \big\rangle.$$

We apply Lemma 6.10 with $\mathcal{A} = (b\lambda_j/2) B_j$, with the drift $\phi = b\lambda_j m_j B_j \nabla F$, which is bounded by (a) and Lipschitz in x since $D^2 F$ is bounded on each piece, and with the last

term as Y. Using $0 \le m_j \le 1$ and $|\nabla \partial_i F| \le \sqrt{\ell}\,\Theta_2$, we find that $|\mathrm{Y}| \le \ell b \lambda_j |B_j| \Theta_2^2$, and Lemma 6.10 yields

$$\Theta_2(\tau) \le \Xi(\tau), \qquad \Xi(\tau) = \|D^2 f\|_\infty + \ell b \int_\tau^N \lambda(\tau')|B(\tau')|\,\Theta_2(\tau')^2 \,\mathrm{d}\tau'.$$

If $\|D^2 f\|_\infty = 0$, then $\Xi(\tau) \le C \int_\tau^N \Xi(\tau')\,\mathrm{d}\tau'$, and Gronwall's inequality implies that $\Xi = 0$. Otherwise Ξ is positive and absolutely continuous, and $-\Xi' = \ell b \lambda |B| \Theta_2^2 \le \ell b \lambda |B|\,\Xi^2$ almost everywhere. Then $(1/\Xi)' \le \ell b \lambda |B|$, and integrating from τ to N, we obtain

$$\frac{1}{\Xi(\tau)} \ge \frac{1}{\|D^2 f\|_\infty} - \ell b \sum_j \lambda_j |B_j| \ge \frac{1}{2\|D^2 f\|_\infty},$$

where we used the hypothesis of (b) in the last step. This proves (b).

For (c), let $|\iota| = r \ge 3$. By the Leibniz rule,

$$-\partial_\tau D^\iota F = \frac{b\lambda_j}{2} B_j : D^2 D^\iota F + b\lambda_j m_j \big\langle B_j \nabla F, \nabla D^\iota F \big\rangle + \frac{b\lambda_j m_j}{2} \sum_{0<\iota'<\iota} \binom{\iota}{\iota'} \big\langle \nabla D^{\iota'} F, B_j \nabla D^{\iota-\iota'} F \big\rangle,$$

where $0 < \iota' < \iota$ means $\iota' \le \iota$ componentwise with $\iota' \notin \{0, \iota\}$. The two factors in the last sum have orders $|\iota'| + 1$ and $r - |\iota'| + 1$, both between 2 and r. In the terms with $|\iota'| \in \{1, r-1\}$, one factor has order r and the other order 2; the remaining terms involve only orders between 3 and $r-1$. By Lemma 6.10, applied as before with the last sum as the source, we have

$$\Theta_r(\tau) \le \|D^r f\|_\infty + C_{\ell,r} b \int_\tau^N \lambda(\tau')|B(\tau')| \Big\{ \Theta_2 \Theta_r + \sum_{q=3}^{r-1} \Theta_q \Theta_{r+2-q} \Big\}(\tau')\,\mathrm{d}\tau'.$$

By (b) and induction on r, the sum over q is bounded. Gronwall's inequality then bounds Θ_r, with a constant of the stated type. This proves (c). □

Proof of Lemma 6.2. We begin with (a). The gradient bound is Lemma 6.11(a). By Lemma 4.2(a), each integration step is monotone and commutes with the addition of constants, which implies that it does not increase the supremum of the difference of two functions. Composing the steps, we find that $\sup|F - F'| \le \sup|f - f'|$. For (6.3), we refine the partitions so that both m_j and $\widetilde{m}_j$ are constant on each piece; by Lemma 4.2(b) this does not change F or $\widetilde{F}$. Because B_j is symmetric, we have

$$m_j \langle \nabla F, B_j \nabla F \rangle - \widetilde{m}_j \langle \nabla \widetilde{F}, B_j \nabla \widetilde{F} \rangle = m_j \big\langle B_j (\nabla F + \nabla \widetilde{F}), \nabla (F - \widetilde{F}) \big\rangle + (m_j - \widetilde{m}_j) \langle \nabla \widetilde{F}, B_j \nabla \widetilde{F} \rangle.$$

Then $\mathcal{D} = F - \widetilde{F}$ satisfies, on stage j,

$$-\partial_\tau \mathcal{D} = \frac{b\lambda_j}{2} B_j : D^2 \mathcal{D} + \frac{b\lambda_j m_j}{2} \big\langle B_j(\nabla F + \nabla \widetilde{F}), \nabla \mathcal{D} \big\rangle + \frac{b\lambda_j}{2} (m_j - \widetilde{m}_j) \langle \nabla \widetilde{F}, B_j \nabla \widetilde{F} \rangle,$$

with $\mathcal{D}(N, \cdot) = 0$. By Lemma 6.9(a) and (b), $\mathcal{D}$ is of class $C^{1,2}$ on each piece, $\nabla \mathcal{D}$ is bounded, and the drift is bounded and Lipschitz in x, since the Hessians of F and $\widetilde{F}$ are bounded on each piece. The source is bounded by $(b/2)\lambda_j |B_j| |m_j - \widetilde{m}_j| L_f^2$, by the gradient bound. Then Lemma 6.10 implies (6.3).

For (b), by (6.3), we see that the step flows $F^{(l)}$ of any sequence with $\int_0^1 |m_j^{(l)} - m_j| \to 0$ form a Cauchy sequence for uniform convergence. Interlacing two such sequences gives another one, and we conclude that the limit does not depend on the sequence. The estimate (6.3) and the bound on $\sup|F - F'|$ pass to uniform limits. For the gradient, let $e_{l,l'} = \sup_{\tau,x} |F^{(l)} - F^{(l')}|$, let v be a unit vector, and let $a > 0$. By Taylor's formula and the Hessian bound of Lemma 6.11(a), which holds for every l, the difference quotient $(F^{(l)}(\tau, x + av) - F^{(l)}(\tau, x))/a$ differs from $\partial_v F^{(l)}(\tau, x)$ by at most $(K_f + L_f^2)a/2$. Comparing the difference quotients of $F^{(l)}$ and $F^{(l')}$, we find that

$$\sup_{\tau,x} |\partial_v F^{(l)}(\tau, x) - \partial_v F^{(l')}(\tau, x)| \le \frac{2e_{l,l'}}{a} + (K_f + L_f^2)a.$$

Choosing $a = e_{l,l'}^{1/2}$, we see that the gradients of the step flows are uniformly Cauchy. Their uniform limit is continuous, and by the fundamental theorem of calculus it is the gradient of F, which therefore satisfies $|\nabla F| \le L_f$.

For (c), let $F^{(l)}$ be step flows of stages $1, \dots, N$ with masses $m_j^{(l)}$ as in (b), and let $f_{\text{in}}^{(l)} = F^{(l)}(N', \cdot)$, which is the output of the step flow of stages $N' + 1, \dots, N$. Let $\bar{F}^{(l)}$ and $\bar{F}$ be the step flow of stages $1, \dots, N'$ with masses $m_j^{(l)}$ and terminal function f_{in}, and its limit. On $[0, N']$, the function $F^{(l)}$ is the step flow of stages $1, \dots, N'$ with terminal function $f_{\text{in}}^{(l)}$. By part (a), we have

$$\sup|F^{(l)} - \bar{F}| \le \sup|f_{\text{in}}^{(l)} - f_{\text{in}}| + \sup|\bar{F}^{(l)} - \bar{F}|.$$

Both terms tend to zero, by (b) for the stages $N' + 1, \dots, N$ and for the stages $1, \dots, N'$. This proves (c).

For (d), we first note that the hypothesis of Lemma 6.11(b) does not involve the masses and holds for every step flow. Using Lemma 6.11, we see that all spatial derivatives of the step flows are bounded by constants of the type stated there, uniformly in l. Let l, l' be two indices, refine the partitions so that $m_j^{(l)}$ and $m_j^{(l')}$ are constant on each piece, and let $\mathcal{D} = F^{(l)} - F^{(l')}$. Then $\mathcal{D}$ satisfies the equation in the proof of (a), with $F = F^{(l)}$, $\widetilde{F} = F^{(l')}$,

and the masses $m_j^{(l)}$, $m_j^{(l')}$ in place of m_j, $\widetilde{m}_j$. For $\iota \neq 0$, we apply D^ι to that equation and use the Leibniz rule. We find that

$$-\partial_\tau D^\iota \mathcal{D} = \frac{b\lambda_j}{2} B_j : D^2 D^\iota \mathcal{D} + \frac{b\lambda_j m_j^{(l)}}{2} \big\langle B_j(\nabla F^{(l)} + \nabla F^{(l')}), \nabla D^\iota \mathcal{D} \big\rangle + \mathrm{Y}_\iota,$$

where Y_ι is the sum of the terms $(b\lambda_j m_j^{(l)}/2)\binom{\iota}{\iota'}\langle B_j D^{\iota'} \nabla (F^{(l)} + F^{(l')}), \nabla D^{\iota-\iota'} \mathcal{D}\rangle$ with $0 \neq \iota' \leq \iota$ and of $(b\lambda_j/2)(m_j^{(l)} - m_j^{(l')}) D^\iota \langle \nabla F^{(l')}, B_j \nabla F^{(l')}\rangle$. The terms with $|\iota'| = 1$ are linear in the derivatives of $\mathcal{D}$ of order $|\iota|$, with coefficients bounded by $C\lambda_j |B_j|$; the terms with $|\iota'| \geq 2$ involve derivatives of $\mathcal{D}$ of order less than $|\iota|$; and the last term is bounded by $C\lambda_j |B_j| |m_j^{(l)} - m_j^{(l')}|$, with C depending on the derivatives of $F^{(l')}$ of order at most $|\iota| + 1$. By Lemma 6.10, induction on $|\iota|$, and Gronwall's inequality, we have

$$\sup_{\tau,x} |D^\iota \mathcal{D}(\tau, x)| \leq C \sum_j \lambda_j \int_0^1 |m_j^{(l)} - m_j^{(l')}| \, \mathrm{d}\tau.$$

It follows that every spatial derivative of the step flows converges uniformly, and the limit F has continuous spatial derivatives of all orders, which are the limits of those of the step flows. Letting $l' \to \infty$ in the last display gives (6.4), and the Hessian bound follows from Lemma 6.11(b).

Now let each m_j be continuous, and choose the step flows with the sampled masses $m_j^{\mathfrak{p}_l}$ for partitions with $|\mathfrak{p}_l| \to 0$; then $m_j^{\mathfrak{p}_l} \to m_j$ pointwise and in L^1. The step flows are continuous in τ and continuously differentiable on each piece, and, for $j - 1 \leq \tau \leq \tau_1 \leq j$, they satisfy

$$F^{(l)}(\tau, x) = F^{(l)}(\tau_1, x) + \int_\tau^{\tau_1} \frac{b\lambda_j}{2} \Big\{ B_j : D^2 F^{(l)} + m_j^{\mathfrak{p}_l} (\tau' - j + 1) \big\langle \nabla F^{(l)}, B_j \nabla F^{(l)} \big\rangle \Big\} (\tau', x) \, \mathrm{d}\tau'.$$

By bounded convergence, the same identity holds for F and m_j. Its integrand is continuous in τ', since ∇F and $D^2 F$ are uniform limits of continuous functions and m_j is continuous. This shows that F is continuously differentiable in τ on $[j-1, j]$ and solves (6.1). For uniqueness, let F' be another solution with the stated properties. The difference $\mathcal{D} = F - F'$ satisfies the equation in the proof of (a), with $\widetilde{F} = F'$, $\widetilde{m}_j = m_j$, and $\mathcal{D}(N, \cdot) = 0$. Its drift is bounded and Lipschitz in x, since the first and second derivatives of F and F' are bounded, and $\mathcal{D}$ and $\nabla \mathcal{D}$ have linear growth. Then Lemma 6.10, with $\mathrm{Y} = 0$, implies that $\mathcal{D} = 0$. □

6.4 Parameter derivatives

We now prove Proposition 6.3. Throughout this section we assume (H1)–(H5), and on stage j we abbreviate

$$\mathcal{A}_j = \frac{bh}{2}\kappa_j B_j, \qquad \mathcal{V}_j = \frac{bh}{2}\kappa_j m_j B_j,$$

so that (6.1) reads $-\partial_\tau F = \mathcal{A}_j : D^2F + \langle \nabla F, \mathcal{V}_j \nabla F\rangle$. Both matrices are symmetric and depend on (h,k), and $\mathcal{V}_j$ also depends on τ through the mass. Their derivatives $\partial^{\gamma'}\mathcal{A}_j = (b/2)\partial^{\gamma'}(h\kappa_j B_j)$ and $\partial^{\gamma'}\mathcal{V}_j$ are bounded by the suprema listed in Proposition 6.3(ii). Note that $\partial_h \mathcal{A}_j$ does not vanish at $h = 0$; these derivatives are bounded but not small. The proof has two parts. In Lemma 6.12, we prove the bounds for the step flows with sampled masses, with constants that do not depend on the partitions. We then pass to the limit.

Lemma 6.12 (Uniform bounds for step flows)**.** *Let $\mathfrak{p}$ be partitions, and let $F^{\mathfrak{p}}(\tau,h,k,x)$ be the step flow with the sampled masses $m_j^{\mathfrak{p}}(\cdot,h,k)$ and terminal function $f(h,k,\cdot)$, for $(h,k)\in\mathcal{U}_0$. Then $F^{\mathfrak{p}}$ is infinitely differentiable in (h,k,x), its derivatives $\partial^\gamma D^\iota F^{\mathfrak{p}}$ are continuous in (τ,h,k,x) and, on each closed piece, continuously differentiable in τ, with ∂_τ commuting with $\partial^\gamma D^\iota$. For every $(\gamma,\iota)\neq(0,0)$, the function $\partial^\gamma D^\iota F^{\mathfrak{p}}$ is bounded on $[0,N]\times\mathcal{U}_0\times\mathbb{R}^\ell$ by a constant C of the type stated in Proposition 6.3(ii), which does not depend on $\mathfrak{p}$.*

Proof. On the piece (j,i), the step (6.2) has mass $m_j(\tau_{j,i-1},h,k)$ and covariance $t\,bh\kappa_j(h,k)B_j(h,k)$, with $t = j-1+\tau_{j,i}-\tau \in [0,1]$. Both are smooth functions of (h,k) on the compact convex set $\mathcal{U}_0$, and the covariance is positive semidefinite for $h\geq 0$. Applying Lemma 6.9(c) to the pieces one after another, from the innermost, we obtain the regularity statements. The commutation of ∂_τ with $\partial^\gamma D^\iota$ holds because $F^{\mathfrak{p}}$ is infinitely differentiable in (τ,h,k,x) on each closed piece. It remains to prove the bounds, which we do by induction. We drop the superscript $\mathfrak{p}$ and write $F_\gamma = \partial^\gamma F$ and

$$\Theta_{\gamma,r}(\tau) = \max_{|\iota|=r}\sup_x |D^\iota F_\gamma(\tau,h,k,x)|,$$

for fixed $(h,k)\in\mathcal{U}_0$; all bounds below are uniform in (h,k).

Step 1: spatial derivatives. For $\gamma = 0$ and $r\geq 1$, the bound on $\Theta_{0,r}$ is Lemma 6.11, which applies since $2\ell b\|D^2 f\|_\infty \sum_j \lambda_j|B_j| \leq 2\ell b\, h_0\Lambda_*\|D^2 f\|_\infty \leq 1$ by (H5). We have $\Theta_{0,2}\leq 2\|D^2 f\|_\infty$.

Step 2: the differentiated equations. Let $\gamma\neq 0$. Applying ∂^γ to (6.1) and separating the terms that contain F_γ, we find that, on each piece of stage j,

$$-\partial_\tau F_\gamma = \mathcal{A}_j : D^2 F_\gamma + 2\langle \mathcal{V}_j\nabla F, \nabla F_\gamma\rangle + \mathrm{Y}_\gamma, \tag{6.13}$$

with

$$\mathrm{Y}_\gamma = \sum_{0\neq\gamma'\le\gamma} \binom{\gamma}{\gamma'} \partial^{\gamma'}\mathcal{A}_j : D^2 F_{\gamma-\gamma'} + \sum_{\substack{\gamma'+\gamma''+\gamma'''=\gamma \\ \gamma''\neq\gamma,\ \gamma'''\neq\gamma}} \frac{\gamma!}{\gamma'!\,\gamma''!\,\gamma'''!} \big\langle \nabla F_{\gamma''}, \partial^{\gamma'}\mathcal{V}_j \nabla F_{\gamma'''} \big\rangle, \quad (6.14)$$

where the second sum runs over the triples of multi-indices with $\gamma'+\gamma''+\gamma'''=\gamma$ other than $(0,\gamma,0)$ and $(0,0,\gamma)$; since $\gamma\neq 0$, these are the triples with $\gamma''\neq\gamma$ and $\gamma'''\neq\gamma$. Every $F_{\gamma''}$, $F_{\gamma'''}$, and $F_{\gamma-\gamma'}$ appearing in Y_γ has parameter order strictly less than $|\gamma|$, and $F_0 = F$ enters only through ∇F and D^2F. Applying D^ι with $|\iota| = r$ to (6.13), we obtain

$$-\partial_\tau D^\iota F_\gamma = \mathcal{A}_j : D^2 D^\iota F_\gamma + 2\langle \mathcal{V}_j \nabla F, \nabla D^\iota F_\gamma \rangle + 2 \sum_{0\neq\iota'\le\iota} \binom{\iota}{\iota'} \big\langle \mathcal{V}_j \nabla D^{\iota'} F, \nabla D^{\iota-\iota'} F_\gamma \big\rangle + D^\iota \mathrm{Y}_\gamma, \quad (6.15)$$

with terminal value $D^\iota F_\gamma(N,\cdot) = \partial^\gamma D^\iota f$.

Step 3: derivative bounds by weight. We prove the bound on $\Theta_{\gamma,r}$ by induction on $|\gamma|$, and, for fixed γ, by induction on $r = 0,1,2,\dots$. Table 6.1 lists the terms of (6.15) and the derivatives of F that they contain. Let $\varpi = r + 2|\gamma|$ be the weight of $D^\iota F_\gamma$. We read off from the table that every derivative $\partial^{\gamma^\circ} D^{\iota^\circ} F$ other than the unknown $D^\iota F_\gamma$ has $\gamma^\circ \le \gamma$ and weight $|\iota^\circ| + 2|\gamma^\circ| \le \varpi$. Moreover, either $|\gamma^\circ| < |\gamma|$, or $\gamma^\circ = \gamma$ and $|\iota^\circ| \le r$. The terms with $\gamma^\circ = \gamma$ and $|\iota^\circ| = r$ are those in the second row. They are linear in the derivatives of order r of F_γ, with coefficients bounded by $2\ell r|\mathcal{V}_j|\Theta_{0,2}$. In the fourth row, a parameter derivative is paid for by two spatial derivatives. Indeed, when $\partial^{\gamma'}$ falls on the coefficient $\mathcal{A}_j$ of D^2, the derivative $F_{\gamma-\gamma'}$ appears with $r+2$ spatial derivatives, and its weight $r + 2 + 2(|\gamma| - |\gamma'|)$ is at most ϖ because $|\gamma'| \ge 1$.

In the last row, $\gamma'' = 0$ or $\gamma''' = 0$ is allowed; the corresponding factor is a spatial derivative of F of order at most $r+1 \le \varpi - 1$, bounded in Step 1. By the induction hypothesis and Step 1, we conclude that all terms of (6.15) other than those in the first two rows are bounded by a constant c_0. The coefficients in the second row are bounded by $c_1 = 2\ell r \max_j \sup|\mathcal{V}_j| \cdot 2\|D^2 f\|_\infty$.

Step 4: the estimate. We apply Lemma 6.10 to each $\psi = D^\iota F_\gamma$ with $|\iota| = r$, with $\mathcal{A} = \mathcal{A}_j$ on stage j, the drift $\phi = 2\mathcal{V}_j \nabla F$, and the source formed by all other terms of (6.15). The drift is bounded by Step 1, and it is Lipschitz in x since D^2F is bounded. By the regularity statement, ψ, $\nabla\psi$, and the source are bounded and continuous on each piece. Taking the maximum over $|\iota| = r$, we find

$$\Theta_{\gamma,r}(\tau) \le \max_{|\iota|=r} \sup_x |\partial^\gamma D^\iota f| + \int_\tau^N \big\{ c_1 \Theta_{\gamma,r}(\tau') + c_0 \big\}\, \mathrm{d}\tau',$$

and by Gronwall's inequality we have $\Theta_{\gamma,r} \le (\max_{|\iota|=r} \sup|\partial^\gamma D^\iota f| + N c_0) e^{N c_1}$. For $r = 0$ the second row is absent and $c_1 = 0$. This completes the induction.

Table 6.1. The terms of (6.15), for $|\iota| = r$ and $\varpi = r + 2|\gamma|$. The second column lists the derivatives of F that each term contains, and the third says how they are controlled.

Term	Derivatives of F	Control
$\mathcal{A}_j : D^2 D^\iota F_\gamma$, $2\langle \mathcal{V}_j \nabla F, \nabla D^\iota F_\gamma \rangle$	the unknown $D^\iota F_\gamma$	operator of Lemma 6.10
$\langle \mathcal{V}_j \nabla D^{\iota'} F, \nabla D^{\iota - \iota'} F_\gamma \rangle$, $\lvert\iota'\rvert = 1$	$D^2 F$; derivatives of order r of F_γ (weight ϖ)	Gronwall's inequality
the same, $\lvert\iota'\rvert \geq 2$	$D^{\lvert\iota'\rvert+1} F$ (weight $\leq r+1$); derivatives of order $r - \lvert\iota'\rvert + 1 < r$ of F_γ	induction on r
$\partial^{\gamma'} \mathcal{A}_j : D^2 D^\iota F_{\gamma-\gamma'}$, $\gamma' \neq 0$	order $r+2$ of $F_{\gamma-\gamma'}$, weight $\leq \varpi$	induction on $\lvert\gamma\rvert$
$\langle D^{\iota_1} \nabla F_{\gamma''}, \partial^{\gamma'} \mathcal{V}_j D^{\iota_2} \nabla F_{\gamma'''} \rangle$, $\iota_1 + \iota_2 = \iota$	orders $\lvert\iota_1\rvert + 1$, $\lvert\iota_2\rvert + 1 \leq r+1$ of $F_{\gamma''}$, $F_{\gamma'''}$ with $\lvert\gamma''\rvert, \lvert\gamma'''\rvert < \lvert\gamma\rvert$; weights $< \varpi$	induction on $\lvert\gamma\rvert$

The constant obtained for (γ, r) involves the bounds for the pairs (γ°, r°) listed in Table 6.1, the bounds on $\partial^{\gamma'} \mathcal{A}_j$ and $\partial^{\gamma'} \mathcal{V}_j$ for $\gamma' \leq \gamma$, and the terminal values. Following the induction back, it depends only on ℓ, N, b, h_1, γ, r, and the suprema of $|\partial^{\gamma'}(\kappa_j B_j)|$, $|\partial^{\gamma'}(\kappa_j m_j B_j)|$, and $|\partial^{\gamma'} D^{\iota'} f|$ over $\gamma' \leq \gamma$ and $0 < |\iota'| + 2|\gamma'| \leq \varpi$, as claimed. None of them depends on $\mathfrak{p}$, since the sampled masses have the same values and (h,k)-derivatives as m_j at the sampling points. □

Proof of Proposition 6.3. Let $\mathfrak{p}_l$ be partitions with $|\mathfrak{p}_l| \to 0$, and let $F^{\mathfrak{p}_l}$ be the step flows with the sampled masses. The proof has three steps: convergence of the values, convergence of the derivatives, and the equations.

Step 1: values. By (H3), m_j is uniformly continuous on $[0,1] \times \mathcal{U}$, and we have $\varepsilon_l = \sup_{\mathcal{U}} \sum_j \int_0^1 |m_j^{\mathfrak{p}_l} - m_j| \, \mathrm{d}\tau \to 0$. Then Lemma 6.2(a) and (b), applied at each (h,k) with $\sum_j \lambda_j |B_j| \leq h_0 \Lambda_*$, give

$$\sup_{[0,N] \times \mathcal{U}_0 \times \mathbb{R}^\ell} |F^{\mathfrak{p}_l} - F| \leq \frac{b}{2} \sup |\nabla_x f|^2 \, h_0 \Lambda_* \, \varepsilon_l \longrightarrow 0.$$

Step 2: derivatives. Fix $(\gamma, \iota) \neq (0,0)$. By Lemma 6.12, the functions $\partial^\gamma D^\iota F^{\mathfrak{p}_l}$ are bounded uniformly in l, and so are their first derivatives in x and in (h,k), whose weights are larger by one and by two, respectively. Their derivative in τ is given on each piece by (6.15), and it is also bounded uniformly in l, since it involves derivatives of weight at most $|\iota| + 2|\gamma| + 2$. Since $\partial^\gamma D^\iota F^{\mathfrak{p}_l}$ is also continuous in τ, the family $\{\partial^\gamma D^\iota F^{\mathfrak{p}_l}\}_l$ is uniformly bounded and uniformly Lipschitz on $[0,N] \times \mathcal{U}_0 \times \mathbb{R}^\ell$. By the Arzelà–Ascoli theorem and a diagonal argument over the compact sets $[0,N] \times \mathcal{U}_0 \times \{|x| \leq R\}$, every subsequence has

a further subsequence along which, for all (γ,ι) at once, $\partial^\gamma D^\iota F^{\mathfrak{p}_l}$ converges uniformly on compact sets to a continuous limit $L_{\gamma,\iota}$. We claim that $L_{\gamma,\iota}=\partial^\gamma D^\iota F$, so that the limits do not depend on the subsequence and the whole sequence converges. Indeed, $L_{0,0}=F$ by Step 1. If $\varphi_l\to\varphi$ and $\partial_a\varphi_l\to\psi$ uniformly on compact sets, where ∂_a is a derivative in h, in k, or in a coordinate of x, then letting $l\to\infty$ in $\varphi_l(y+\eta e_a)-\varphi_l(y)=\int_0^\eta \partial_a\varphi_l(y+\eta' e_a)\,\mathrm{d}\eta'$ shows that $\partial_a\varphi=\psi$; at boundary points of $\mathcal{U}_0$, such as $h=0$, we take η of the sign that keeps $y+\eta e_a$ in the domain. Induction on $|\gamma|+|\iota|$ proves the claim. This proves (iv), the existence and continuity of all derivatives in (ii), and the bounds in (ii), which pass to the limit. For the linear growth, note that $|F(\tau,h,k,x)-F(\tau,h,k,0)|\le \sup|\nabla_x f||x|$, that $|F(\tau,h,k,0)-f(h,k,0)|\le N\sup|\partial_\tau F|$ is bounded by (6.1) and the bounds on ∇F and D^2F, and that $\sup_{\mathcal{U}}|f(h,k,0)|$ is finite, since $f(\cdot,\cdot,0)$ is continuous on the compact set $\mathcal{U}$. The constant obtained depends on this supremum in addition to the derivative bounds.

Step 3: the equations. Part (i) is Lemma 6.2(d), whose hypothesis holds at each $(h,k)\in\mathcal{U}_0$ by (H5). For (iii), on stage j and for $j-1\le\tau\le\tau_1\le j$, the step flows satisfy

$$F_\gamma^{\mathfrak{p}_l}(\tau)=F_\gamma^{\mathfrak{p}_l}(\tau_1)+\int_\tau^{\tau_1}\big\{\mathcal{A}_j:D^2F_\gamma^{\mathfrak{p}_l}+2\langle\mathcal{V}_j^{\mathfrak{p}_l}\nabla F^{\mathfrak{p}_l},\nabla F_\gamma^{\mathfrak{p}_l}\rangle+\Upsilon_\gamma^{\mathfrak{p}_l}\big\}(\tau')\,\mathrm{d}\tau',$$

where $F_\gamma^{\mathfrak{p}_l}=\partial^\gamma F^{\mathfrak{p}_l}$ and where $\mathcal{V}_j^{\mathfrak{p}_l}$ and $\Upsilon_\gamma^{\mathfrak{p}_l}$ are formed with the sampled masses. By (H3), the (h,k)-derivatives of $m_j^{\mathfrak{p}_l}$ converge pointwise to those of m_j, and by (iv) the derivatives of $F^{\mathfrak{p}_l}$ converge pointwise. All terms are bounded, and by bounded convergence the same identity holds for $F_\gamma=\partial^\gamma F$, with the masses m_j. Since its integrand is continuous in τ', the function F_γ is continuously differentiable in τ on $[j-1,j]$ and satisfies (6.13), which is the equation obtained by applying ∂^γ to (6.1). This proves (iii). □

6.5 Response to a change of the terminal function

We now compare a hierarchy with terminal function f and the same hierarchy with terminal function $f+V$, for a bounded function V. The first-order change is the expectation of V under the path law of the unperturbed hierarchy, and the error is quadratic in $\|V\|_\infty$, with a constant that does not depend on the number of levels. For a hierarchy with M levels, let $\Sigma_1,\dots,\Sigma_M\succeq 0$ be the covariances and $m_1,\dots,m_M\in[0,1]$ the masses, listed from the root; the masses need not be ordered. For a function f of linear growth, the partial values are $\Psi_M^f=f$ and $\Psi_{j-1}^f=T_{m_j,\Sigma_j}\Psi_j^f$ for $j=M,\dots,1$, as in (4.13), and we write $\mathcal{H}f=\Psi_0^f$. For $z_0\in\mathbb{R}^\ell$, we let $\mathbb{E}_f$ denote expectation for the Markov chain $z_0,z_1,\dots,z_M$ in which z_j has law $\mathcal{K}_{m_j,\Sigma_j,\Psi_j^f}(z_{j-1},\cdot)$ given $z_0,\dots,z_{j-1}$. This is the path law of Definition 4.14(c), started at z_0 and without the terminal draw of spins.

Lemma 6.13 (Response of a hierarchy to its terminal function)**.** *Let f have linear growth, and let $V\colon \mathbb{R}^\ell \to \mathbb{R}$ be bounded and measurable. Then, for every $z_0 \in \mathbb{R}^\ell$,*

$$\bigl|\mathcal{H}(f+V)(z_0) - \mathcal{H}f(z_0) - \mathbb{E}_f V(z_M)\bigr| \le \frac{1}{2}\|V\|_\infty^2.$$

The bound does not depend on the number M of levels or on the covariances.

The argument is the martingale argument that the proof of [77, Lemma 3.1] uses to show that $\chi \le 1 - u^2$, and the differentiation of a hierarchy behind Proposition 4.18, namely [77, Lemma 2.2]; we apply them to the terminal function instead of the field.

Proof. For $0 \le t \le 1$, let $\Psi_j^t = \Psi_j^{f+tV}$ be the partial values with terminal function $f + tV$. Using the monotonicity and the translation identity of Lemma 4.2(a), we have $|\Psi_j^t - \Psi_j^0| \le t\|V\|_\infty$, and all integrals below are finite. Let $\mathbb{E}_j^t$ and Var_j^t denote the expectation and the variance under the kernel $\mathcal{K}_{m_j,\Sigma_j,\Psi_j^t}(z,\cdot)$ for a fixed starting point z. We claim that $t \mapsto \Psi_{j-1}^t(z)$ is twice continuously differentiable, with

$$\partial_t \Psi_{j-1}^t(z) = \mathbb{E}_j^t\bigl[\partial_t \Psi_j^t\bigr], \qquad \partial_t^2 \Psi_{j-1}^t(z) = \mathbb{E}_j^t\bigl[\partial_t^2 \Psi_j^t\bigr] + m_j \operatorname{Var}_j^t\bigl(\partial_t \Psi_j^t\bigr), \tag{6.16}$$

where the expectation and the variance act on the function $y \mapsto \partial_t \Psi_j^t(y)$ or $\partial_t^2 \Psi_j^t(y)$. Indeed, let $\mathcal{Z}_t$ be random variables, twice differentiable in t with $|\partial_t \mathcal{Z}_t| + |\partial_t^2 \mathcal{Z}_t|$ bounded and $\mathbb{E}e^{|\mathcal{Z}_0|} < \infty$, and let $\mathfrak{a}(t) = m^{-1}\log \mathbb{E}e^{m\mathcal{Z}_t}$ for some $m \in (0,1]$. Differentiation under the expectation gives $\mathfrak{a}' = \mathbb{E}[\partial_t \mathcal{Z}_t\, e^{m\mathcal{Z}_t}]/\mathbb{E}e^{m\mathcal{Z}_t}$ and

$$\mathfrak{a}'' = \frac{\mathbb{E}[\partial_t^2 \mathcal{Z}_t\, e^{m\mathcal{Z}_t}]}{\mathbb{E}e^{m\mathcal{Z}_t}} + m\Bigl(\frac{\mathbb{E}[(\partial_t \mathcal{Z}_t)^2\, e^{m\mathcal{Z}_t}]}{\mathbb{E}e^{m\mathcal{Z}_t}} - \Bigl(\frac{\mathbb{E}[\partial_t \mathcal{Z}_t\, e^{m\mathcal{Z}_t}]}{\mathbb{E}e^{m\mathcal{Z}_t}}\Bigr)^2\Bigr).$$

For $m = 0$ and $\mathfrak{a}(t) = \mathbb{E}\mathcal{Z}_t$, the same formulas hold without the variance term. We apply this with $\mathcal{Z}_t = \Psi_j^t(z + \Sigma_j^{1/2}\mathbf{g})$ and $m = m_j$; the weights $e^{m_j\mathcal{Z}_t}/\mathbb{E}e^{m_j\mathcal{Z}_t}$ are the densities of the kernel $\mathcal{K}_{m_j,\Sigma_j,\Psi_j^t}(z,\cdot)$. Starting from $\partial_t \Psi_M^t = V$ and $\partial_t^2 \Psi_M^t = 0$, induction on $M - j$ shows that the derivatives are bounded and proves (6.16).

Now let $z_0, \dots, z_M$ be the chain of the lemma for the terminal function $f + tV$, and let $\mathcal{M}_j = \partial_t \Psi_j^t(z_j)$. By the first identity in (6.16) and the Markov property, we have $\mathbb{E}^t[\mathcal{M}_j \mid z_0, \dots, z_{j-1}] = \mathcal{M}_{j-1}$. This implies that $(\mathcal{M}_j)$ is a martingale, with $\mathcal{M}_M = V(z_M)$ and $\mathcal{M}_0 = \partial_t \mathcal{H}(f+tV)(z_0)$. Iterating the second identity in (6.16) from $\partial_t^2 \Psi_M^t = 0$, we obtain

$$\partial_t^2 \mathcal{H}(f+tV)(z_0) = \sum_{j=1}^{M} m_j\, \mathbb{E}^t\bigl(\mathcal{M}_j - \mathcal{M}_{j-1}\bigr)^2.$$

Since $0 \le m_j \le 1$ and the increments of a martingale are orthogonal,

$$0 \le \partial_t^2 \mathcal{H}(f+tV)(z_0) \le \sum_{j=1}^{M} \mathbb{E}^t\bigl(\mathcal{M}_j - \mathcal{M}_{j-1}\bigr)^2 = \mathbb{E}^t V(z_M)^2 - \bigl(\mathbb{E}^t V(z_M)\bigr)^2 \le \|V\|_\infty^2.$$

At $t = 0$, we have $\partial_t \mathcal{H}(f + tV)(z_0) = \mathcal{M}_0 = \mathbb{E}_f V(z_M)$. By Taylor's formula with integral remainder,

$$\mathcal{H}(f + V)(z_0) - \mathcal{H}f(z_0) - \mathbb{E}_f V(z_M) = \int_0^1 (1 - t)\, \partial_t^2 \mathcal{H}(f + tV)(z_0)\, \mathrm{d}t,$$

which lies in $[0, \|V\|_\infty^2/2]$. This completes the proof. □

We next apply Lemma 6.13 to the scalar hierarchy that integrates a block of $\mathfrak{n}$ identical coordinates over a time interval $[0, t_\star]$, and pass to a general distribution function. For a distribution function α, an integer $\mathfrak{n} \geq 1$, a time $t_\star \in (0, 1]$, and $f \in \mathcal{C}^2(\mathbb{R})$, let $\mathcal{H}^\alpha_{t_\star,\mathfrak{n}} f$ denote the value at $z = 0$ of the flow of Lemma 6.2(b) with $\ell = 1$, $N = 1$, $B_1 = 1$, $\lambda_1 = t_\star$, mass $m_1(\tau) = \alpha(t_\star \tau)/\mathfrak{n}$, and terminal function f. If α_n are step distribution functions with $\int_0^1 |\alpha_n - \alpha|\, \mathrm{d}t \to 0$, then $\mathcal{H}^\alpha_{t_\star,\mathfrak{n}} f$ is the limit of the values at $z = 0$ of the step flows with masses $\alpha_n(t_\star \tau)/\mathfrak{n}$. In the scalar time $t = t_\star \tau$, these step flows solve $-\partial_t F = (b/2)\{\partial_{zz} F + (\alpha_n(t)/\mathfrak{n})(\partial_z F)^2\}$ on each interval where α_n is constant, with $F(t_\star, \cdot) = f$. For a general α, we do not assert that the limiting flow is differentiable in time at the jumps of α. Recall the function $\Phi = \Phi_\alpha$ and the diffusion Y of (4.6).

Corollary 6.14 (Response of the common part of a hierarchy)**.** *Let α be a distribution function, let $\mathfrak{n} \geq 1$ be an integer, and let $t_\star \in (0, 1]$. Then $\mathcal{H}^\alpha_{t_\star,\mathfrak{n}}(\mathfrak{n}\Phi_\alpha(t_\star, \cdot)) = \mathfrak{n}\Phi_\alpha(0, 0)$. Moreover, let $V \in C^2(\mathbb{R})$ with V, V', and V'' bounded, and suppose that $\|V\|_\infty \leq \varepsilon$. Then*

$$\left|\mathcal{H}^\alpha_{t_\star,\mathfrak{n}}\big(\mathfrak{n}\Phi_\alpha(t_\star, \cdot) + V\big) - \mathfrak{n}\Phi_\alpha(0, 0) - \mathbb{E}V(Y_{t_\star})\right| \leq \frac{\varepsilon^2}{2}.$$

In Section 7, we apply the corollary with $\mathfrak{n} = 3$, with $t_\star$ the left endpoint of the window, and with V the difference between the restriction of the local value to the diagonal and $3\Phi_\alpha(t_\star, \cdot)$, for which $\varepsilon = Ch^2k^2$. The initial segment of the hierarchy then contributes $\mathbb{E}V(Y_{t_\star}) + O(h^4k^4)$. Note that the law of $Y_{t_\star}$ does not depend on the perturbation.

Proof. Let α_n be step distribution functions with $t_\star$ among their breakpoints and $\int_0^1 |\alpha_n - \alpha|\, \mathrm{d}t \to 0$; for instance, let α_n be equal to $\alpha(t_{n,i-1})$ on $[t_{n,i-1}, t_{n,i})$, where the $t_{n,i}$ are the points $i2^{-n}$ together with $t_\star$. Let $\Phi_n = \Phi_{\alpha_n}$, let Y^n be the diffusion of α_n, driven by the same Brownian motion as Y, and let $\mathcal{H}_n$ denote the step flow with masses $\alpha_n(t_\star \tau)/\mathfrak{n}$, that is, the finite hierarchy whose levels are the intervals between consecutive breakpoints of α_n in $[0, t_\star]$.

Step 1: the step hierarchies. By (4.3) and Lemma 4.2(d), the partial values of $\mathcal{H}_n$ with terminal function $\mathfrak{n}\Phi_n(t_\star, \cdot)$ are $\mathfrak{n}\Phi_n(t_{n,i}, \cdot)$, and in particular $\mathcal{H}_n(\mathfrak{n}\Phi_n(t_\star, \cdot))(0) = \mathfrak{n}\Phi_n(0, 0)$. The level from $t_{n,i-1}$ to $t_{n,i}$ has mass $\alpha_n(t_{n,i-1})/\mathfrak{n}$ and partial value $\mathfrak{n}\Phi_n(t_{n,i}, \cdot)$. Its kernel has the density $\exp\{\alpha_n(t_{n,i-1})(\Phi_n(t_{n,i}, y) - \Phi_n(t_{n,i-1}, z))\}$ with respect to

$N(z, b(t_{n,i} - t_{n,i-1}))$. This is the kernel of the scalar recursion of α_n on the same interval. By Proposition 4.3(iii), the path law of this hierarchy, started at 0, is the law of $(Y^n_{t_{n,i}})_{t_{n,i} \le t_\star}$. Then Lemma 6.13, with $f = \mathfrak{n}\Phi_n(t_\star, \cdot)$ and $z_0 = 0$, gives

$$\big|\mathcal{H}_n\big(\mathfrak{n}\Phi_n(t_\star, \cdot) + V\big)(0) - \mathfrak{n}\Phi_n(0,0) - \mathbb{E}V(Y^n_{t_\star})\big| \le \frac{\varepsilon^2}{2}. \tag{6.17}$$

Step 2: the limit. By Proposition 4.3(i), $|\partial_z \Phi_n| \le 1$ and $0 < \partial_{zz}\Phi_n \le 1$, and the same holds for Φ_α; hence the terminal functions below belong to $\mathcal{C}^2(\mathbb{R})$, with gradients bounded by $\mathfrak{n} + \|V'\|_\infty$. Using Lemma 6.2(a), we find that

$$|\mathcal{H}_n(\mathfrak{n}\Phi_n(t_\star, \cdot) + V) - \mathcal{H}_n(\mathfrak{n}\Phi_\alpha(t_\star, \cdot) + V)| \le \mathfrak{n} \sup_z |\Phi_n(t_\star, z) - \Phi_\alpha(t_\star, z)|,$$

which tends to zero by Proposition 4.3(ii). By Lemma 6.2(b), $\mathcal{H}_n(\mathfrak{n}\Phi_\alpha(t_\star, \cdot) + V)(0) \to \mathcal{H}^\alpha_{t_\star, \mathfrak{n}}(\mathfrak{n}\Phi_\alpha(t_\star, \cdot) + V)$, since $\int_0^1 |\alpha_n(t_\star \tau) - \alpha(t_\star \tau)| \, \mathrm{d}\tau \to 0$. Also, Proposition 4.3(ii) implies that $\Phi_n(0,0) \to \Phi_\alpha(0,0)$. Finally, we write $u_n = \partial_z \Phi_n$ and $u = \partial_z \Phi_\alpha$. Since $|u_n| \le 1$, $\alpha \le 1$, and $|\partial_z u| \le 1$, we have

$$|\alpha_n(t) u_n(t, y') - \alpha(t) u(t, y)| \le |\alpha_n(t) - \alpha(t)| + \sup_{t,z} |u_n - u| + |y' - y|,$$

and the equations for Y^n and Y give, for $t \le t_\star$,

$$|Y^n_t - Y_t| \le b \int_0^1 |\alpha_n - \alpha| \, \mathrm{d}t' + b \sup |u_n - u| + b \int_0^t |Y^n_{t'} - Y_{t'}| \, \mathrm{d}t'.$$

By Gronwall's inequality and Proposition 4.3(ii), $Y^n_{t_\star} \to Y_{t_\star}$ almost surely, and $\mathbb{E}V(Y^n_{t_\star}) \to \mathbb{E}V(Y_{t_\star})$ by bounded convergence. Taking $n \to \infty$ in (6.17) proves the second claim. The first claim follows in the same way from the identity $\mathcal{H}_n(\mathfrak{n}\Phi_n(t_\star, \cdot))(0) = \mathfrak{n}\Phi_n(0,0)$. □

Remark 6.15 (A common block acts on the diagonal). Let $e \in \mathbb{R}^\ell$, and suppose that stage j of a stage datum has $B_j = ee^\mathsf{T}$. Then the Gaussian increments of that stage are of the form ζe with a centered Gaussian variable ζ, and a point of the line $\{ze : z \in \mathbb{R}\}$ stays on it. It follows that, for $F \in \mathcal{C}^2(\mathbb{R}^\ell)$, the flow of this stage with terminal function F, evaluated at ze, is the scalar flow with $B_j = 1$, the same λ_j and m_j, and terminal function $z \mapsto F(ze)$, evaluated at z. This holds for step flows by Lemma 4.2(e) and passes to the limit by Lemma 6.2(b). In Section 7, the part of the hierarchy before the window has $B = ee^\mathsf{T}$ with $e = (1,1,1)$ and masses $\alpha/3$, and its value is $\mathcal{H}^\alpha_{t_\star, 3}$ applied to the restriction of the local value to the diagonal.

SECTION 7

A three-replica covariance perturbation

In this section, we construct a trial path for the pressure of three configurations whose overlaps are prescribed near a level $r \in (0, q_\beta)$, and we show that its trial value lies below $3P_\beta$ by an amount of third order in the width parameter of the perturbation. The prescribed overlap matrix is $D(r, q, \delta)$ of (4.16), with $q - r$ and δ small. There are two main results. Proposition 7.13 expands the trial value to third order with an explicit polynomial coefficient. Proposition 7.15 turns this expansion into a deficit of order ρ^3, uniformly over the matrices with $\rho \le q - r \le d_0$ and $E\rho \le |\delta| \le 2E\rho$. The deficit is used in Section 8, in the proof of Lemma 8.7, and through it in the static estimate of Section 9.

We use the following earlier results. From Section 4, we use the integration steps of Definition 4.1 and their properties (Lemma 4.2), the hierarchies, penalties and trial values of Definitions 4.14 and 4.15 and (4.14), block factorization (Lemma 4.20), and the matrix interpolation bound Proposition 4.19. From Section 5, we use the bounds on spatial derivatives of Lemma 5.1 and the constant m_β of Corollary 5.5. We also use the regularity of α and Φ on $(0, q_\beta)$ and the stationarity identities (5.15), both from Lemma 5.3. From Section 6, we use the uniform estimates for short compositions of Gaussian flows (Proposition 6.3), the identification of these flows with limits of hierarchies (Lemma 6.2), and the response of a hierarchy to a small change of its terminal function (Lemma 6.13 and Corollary 6.14).

Relation to [77, Section 7]. The construction refines the narrow-annulus comparison of OpenAI's barrier paper [77, Section 7, Steps 1–8]. We keep three of its ingredients. The first is the three-coordinate hierarchy in which the masses are divided by the sizes of the common blocks [77, Section 4.2 and Section 7, Step 1, (7.7)]; OpenAI attributes it to Talagrand's three-replica construction [105, discussion after Proposition 4.4]. The second is the perturbation that adds a Gaussian linear combination of field increments to the field of x, subtracts it from the field of y, and leaves v unchanged [77, Section 7, Step 2, (7.8)–(7.12)]. The third is the source of the loss: the identity $\beta^2 \mathbb{E}\chi(r, Y_r)^2 = 1$, which follows from stationarity, together with the evolution of the susceptibility χ along the scalar diffusion [77, Section 7, Step 8, (7.31)]. In [77], the window has width ρ^8 and the imbalance δ is of order ρ^{20}. The endpoint of the perturbed path differs from the prescribed matrix, and the mismatch is paid through the term $(\beta^2/4)\|D - Q_N\|_F^2$ of Proposition 4.19 [77, (7.12)–(7.13)]. The resulting deficit is of order ρ^{43} [77, (7.32)].

Three parts of the present construction are new. First, the window has width parameter $h = (q - r)/\vartheta$ for a fixed ratio ϑ. The Gaussian contrast has fixed weights and extends past the time at which x and y separate, and a terminal correction of the covariance removes the excess covariance in the antisymmetric direction $\mathbf{e}_x - \mathbf{e}_y$ exactly. The path is positive semidefinite at every stage, its endpoint is $D(r, q, \delta)$, and no mismatch is paid (Section 7.1). Second, we expand the trial value to third order in h. At second order the first stationarity identity cancels the penalty, and at third order we use the second stationarity identity to combine two coefficient polynomials into one (Sections 7.3 to 7.5). The identities themselves are known; their sources are recalled in Section 5. Third, the resulting polynomial gives a deficit of order ρ^3 on a band where $q - r$ and $|\delta|$ are both comparable to ρ (Section 7.6).

Organization. Section 7.1 defines the perturbed path and its trial value and reduces the trial value to a boundary perturbation V on the diagonal. Section 7.2 records the derivative bounds for the local part of the path. In Section 7.3 we compute the Taylor coefficients of V in the width parameter h when the Parisi distribution function is replaced by a constant on the window, using covariance matrices that are not positive semidefinite as formal bookkeeping devices. In Section 7.4 we compute the contribution of the variation of α on the window and prove the pointwise expansion Proposition 7.12. Section 7.5 proves Proposition 7.13, and Section 7.6 proves the uniform deficit Proposition 7.15.

Standing notation. Throughout this section, $\beta > 1$ is fixed and $b = \beta^2$. We write $\alpha = \alpha_\beta$, $\Phi = \Phi_\alpha$, $u = \partial_z \Phi$, $\chi = \partial_z^2 \Phi$ and $w = \partial_z^3 \Phi$, and Y is the scalar diffusion of (4.7). Vectors of $\mathbb{R}^3$ are written in bold, in the coordinate order (v, x, y). We write $\mathbf{e} = (1, 1, 1)^\mathsf{T}$, $\mathbf{e}_v, \mathbf{e}_x, \mathbf{e}_y$ for the unit vectors,

$$L_2 = (\mathbf{e}_x - \mathbf{e}_y)(\mathbf{e}_x - \mathbf{e}_y)^\mathsf{T},$$

and $\mathbf{z} = (z_v, z_x, z_y)$ for a point of $\mathbb{R}^3$. A real number z denotes the point $z\mathbf{e}$ of the diagonal. As in Section 4.4, $U : U' = \sum_{i,l} U_{il} U'_{il}$ and $\|U\|_F^2 = U : U$. For a symmetric 3×3 matrix B and a function F on $\mathbb{R}^3$, we write $B : D^2 F = \sum_{i,l} B_{il} \partial_i \partial_l F$ and $\nabla F \cdot B \nabla F$. We also set $\mathrm{cub}(U) = \sum_{i,l} U_{il}^3$.

We fix a ratio $\vartheta \in (0, 1/2]$, a bound $k_0 > 0$ for the relative imbalance k below, and a compact interval $K \subset (0, q_\beta)$ of left endpoints. With L_ϑ as in (7.1) below, we set $\tau_{\max} = 2 + 4k_0^2 L_\vartheta$, and we fix $h_1 \in (0, 1]$ such that

$$K' = \big[\min K,\ \max K + \tau_{\max} h_1\big] \subset (0, q_\beta).$$

By Lemma 5.3, α is infinitely differentiable on K', the number $\alpha_* = \min_{K'} \alpha$ is positive, α is Lipschitz on K' with constant K_β, and every derivative $\partial_t^i \partial_z^j \Phi$ with $i + j \geq 1$ is bounded on $K' \times \mathbb{R}$. In this section, C denotes a constant that depends only on β, ϑ, k_0 and K (and therefore on h_1 and K'); its value may change from line to line. We make no claim of uniformity as $\vartheta \downarrow 0$. The symbols of the construction are collected in Table 7.1.

Symbol	Meaning	Defined in
ϑ	ratio $(q-r)/h$, fixed in $(0,1/2]$	(7.1)
q_0, h, k	left endpoint, width parameter, relative imbalance	(7.1)
r, q, d, δ	q_0+h, $q_0+(1+\vartheta)h$, ϑh, hk	(7.1)
ω_0, ω_1, L_ϑ	weights of the contrast and its variance	(7.1)
κ_j, $\mathfrak{n}_j$, B_j, B_j°	lengths, block sizes and rates of the stages	Table 7.2
τ_T, T	$2+4k^2L_\vartheta$ and $q_0+h\tau_T$	(7.3)
$\mathcal{T}$, $\mathcal{B}$, V, $\Delta\Pi$	trial value, local block, boundary perturbation, change of the penalty	Section 7.1
Λ_0, Λ_1, Λ_2	coefficient matrices at constant mass	(7.20)
C_χ, C_w, N_ϑ	coefficient polynomials	(7.26), (7.27), (7.15)
P_ϑ, c_j, M	cubic polynomial, its coefficients, scalar factor	Section 7.5

Table 7.1. Notation of Section 7.

7.1 The perturbed covariance path

Parameters and stages

For $q_0 \in (0,1)$, $h \geq 0$ and $k \in \mathbb{R}$, we set

$$\begin{gathered} r = q_0 + h, \qquad q = q_0 + (1+\vartheta)h, \qquad d = q - r = \vartheta h, \qquad \delta = hk, \\ \omega_0 = \frac{1}{2-\vartheta}, \qquad \omega_1 = \frac{1-\vartheta}{2-\vartheta}, \qquad L_\vartheta = \omega_0^2(3-\vartheta) + \omega_1^2 = \frac{\vartheta^2 - 3\vartheta + 4}{(2-\vartheta)^2}. \end{gathered} \tag{7.1}$$

Note that $\omega_0 + \omega_1 = 1$ and $\omega_1 = (1-\vartheta)\omega_0$. The number k is the imbalance relative to the width. We define the vectors

$$\mathbf{c}_0 = \begin{pmatrix} 1 \\ 1+\omega_0 k \\ 1-\omega_0 k \end{pmatrix}, \qquad \mathbf{c}_1 = \begin{pmatrix} 0 \\ 1-\omega_0 k \\ 1+\omega_0 k \end{pmatrix}, \qquad \mathbf{c}_v = \begin{pmatrix} 1 \\ \omega_1 k \\ -\omega_1 k \end{pmatrix},$$
$$\mathbf{c}_x = \begin{pmatrix} 0 \\ 1-\omega_0 k \\ \omega_0 k \end{pmatrix}, \qquad \mathbf{c}_y = \begin{pmatrix} 0 \\ -\omega_0 k \\ 1+\omega_0 k \end{pmatrix},$$

the rates

$$\begin{gathered} B_1 = \mathbf{c}_0\mathbf{c}_0^\intercal, \quad B_2 = \mathbf{c}_1\mathbf{c}_1^\intercal, \quad B_3 = \mathbf{c}_v\mathbf{c}_v^\intercal, \quad B_4 = \mathbf{c}_v\mathbf{c}_v^\intercal + \mathbf{c}_x\mathbf{c}_x^\intercal + \mathbf{c}_y\mathbf{c}_y^\intercal, \quad B_5 = \mathrm{Id} - \tfrac{1}{4}L_2, \\ B_1^\circ = \mathbf{e}\mathbf{e}^\intercal, \quad B_2^\circ = (\mathbf{e}_x+\mathbf{e}_y)(\mathbf{e}_x+\mathbf{e}_y)^\intercal, \quad B_3^\circ = \mathbf{e}_v\mathbf{e}_v^\intercal, \quad B_4^\circ = B_5^\circ = \mathrm{Id}, \end{gathered} \tag{7.2}$$

and the lengths $\kappa_1 = 1$, $\kappa_2 = \kappa_3 = \vartheta$, $\kappa_4 = 1 - \vartheta$ and $\kappa_5 = 4k^2 L_\vartheta$. Note that $B_j = B_j^\circ$ for $j \le 4$ when $k = 0$, and that $\kappa_5 = 0$ when $k = 0$. We also set

$$\tau_T = \kappa_1 + \kappa_2 + \kappa_4 + \kappa_5 = 2 + 4k^2 L_\vartheta, \qquad T = q_0 + h\tau_T. \tag{7.3}$$

The length κ_3 does not enter τ_T, because stages 2 and 3 below have the same scalar interval; $h\tau_T$ is the length of the window $[q_0, T]$.

The *local window* consists of the five stages of Table 7.2, listed from the root to the terminal time; their Gaussian integrations are applied in the reverse order, from stage 5 to stage 1. Stage j has scalar interval $I_j = [q_0 + ho_j,\, q_0 + h(o_j + \kappa_j)]$, with offsets $o_1 = 0$, $o_2 = o_3 = 1$, $o_4 = 1 + \vartheta$ and $o_5 = 2$, and its covariance increment over a scalar time $\mathrm{d}t$ in I_j is $B_j\,\mathrm{d}t$ (before multiplication by b), integrated at mass $\alpha(t)/\mathfrak{n}_j$. Stages 2 and 3 have the same scalar interval $[r, q]$; they are different blocks of the hierarchy, and stage 2 precedes stage 3 along the path, so that the integration of stage 3 is applied first. In the *baseline* window, the rates B_j are replaced by B_j° and everything else is unchanged; in particular, the baseline has the same stage 5, of length κ_5, with the rate Id.

j	stage	scalar interval I_j	length κ_j	mass	rate
1	common	$[q_0, r]$	1	$\alpha(t)/3$	B_1
2	shared	$[r, q]$	ϑ	$\alpha(t)/2$	B_2
3	singleton, first part	$[r, q]$	ϑ	$\alpha(t)$	B_3
4	singleton, second part	$[q, q_0 + 2h]$	$1 - \vartheta$	$\alpha(t)$	B_4
5	terminal correction	$[q_0 + 2h, T]$	$4k^2 L_\vartheta$	$\alpha(t)$	B_5

Table 7.2. The five stages of the local window. The length is the length of I_j divided by h, and the mass of stage j is $\alpha(t)/\mathfrak{n}_j$ with block sizes $\mathfrak{n}_1 = 3$, $\mathfrak{n}_2 = 2$, $\mathfrak{n}_3 = \mathfrak{n}_4 = \mathfrak{n}_5 = 1$.

The *perturbed path* consists of an initial segment on $[0, q_0]$, on which the three coordinates share every increment, with rate $\mathbf{e}\mathbf{e}^\mathsf{T}$ and mass $\alpha(t)/3$; the local window; and a final segment on $[T, 1]$ with rate Id and mass $\alpha(t)$. The *baseline path* is the same with the baseline window. It is the path of Lemma 4.20(b), with the stages split at the scalar times q_0, $q_0 + 2h$ and T.

The contrast. The rates have the following meaning. In this description we assume $h > 0$; at $h = 0$ the window has length zero, and all its covariance increments vanish. Let g_1, g_2, g_v, g_x and g_y be independent centered Gaussian variables with variances bh, $b\vartheta h$, bh, $b(1-\vartheta)h$ and $b(1-\vartheta)h$. In the baseline window, these are the field increments: the three coordinates share g_1 on $[q_0, r]$, the coordinates x and y share g_2 on $[r, q]$, the coordinate v receives g_v on $[r, q_0 + 2h]$ (split between stages 3 and 4), and x and y receive g_x and g_y on $[q, q_0 + 2h]$. We define the contrast

$$\Xi = \frac{1}{h}\big\{\omega_0(g_1 - g_2 - g_x - g_y) + \omega_1 g_v\big\}. \tag{7.4}$$

The perturbation adds $\delta\,\Xi$ to the field of x, subtracts it from the field of y, and leaves the field of v unchanged. Collecting the coefficients of the five Gaussian variables, we find that the perturbed field vector of the window is

$$\mathbf{c}_0 g_1 + \mathbf{c}_1 g_2 + \mathbf{c}_v g_v + \mathbf{c}_x g_x + \mathbf{c}_y g_y.$$

This explains the rates $B_1, \dots, B_4$. By independence,

$$\begin{aligned} \operatorname{Var}\Xi &= \frac{b}{h}\big\{\omega_0^2\big(1+\vartheta+2(1-\vartheta)\big)+\omega_1^2\big\} = \frac{bL_\vartheta}{h}, \\ \operatorname{Cov}(\Xi, g_1+g_2+g_x) &= b\,\omega_0\big(1-\vartheta-(1-\vartheta)\big) = 0, \\ \operatorname{Cov}(\Xi, g_1+g_v) &= b(\omega_0+\omega_1) = b, \end{aligned} \tag{7.5}$$

and the same holds with g_y in place of g_x. These identities imply that the perturbation changes the covariance of the window fields, divided by b, by

$$\delta\big(\mathbf{e}_v(\mathbf{e}_x-\mathbf{e}_y)^{\mathsf{T}} + (\mathbf{e}_x-\mathbf{e}_y)\mathbf{e}_v^{\mathsf{T}}\big) + \frac{\delta^2 L_\vartheta}{h}\,L_2.$$

The first term changes the overlaps of v with x and with y by $+\delta$ and $-\delta$, respectively. The second term is an excess covariance in the antisymmetric direction $\mathbf{e}_x - \mathbf{e}_y$. We have $\delta^2 L_\vartheta/h = hk^2 L_\vartheta = h\kappa_5/4$. Stage 5 uses the rate $\mathrm{Id} - L_2/4$ instead of Id during a time $h\kappa_5$, and it removes this excess exactly. We call stage 5 the terminal correction.

Lemma 7.1 (Identities of the window)**.** *For every $k \in \mathbb{R}$, the following hold.*

(a) $\mathbf{e}\cdot B_j\mathbf{e} = \mathbf{e}\cdot B_j^\circ\mathbf{e}$ *for* $1 \le j \le 5$*; these numbers are* 9, 4, 1, 3 *and* 3.

(b) $\sum_j \kappa_j B_j = \sum_j \kappa_j B_j^\circ + k\big(\mathbf{e}_v(\mathbf{e}_x-\mathbf{e}_y)^{\mathsf{T}} + (\mathbf{e}_x-\mathbf{e}_y)\mathbf{e}_v^{\mathsf{T}}\big)$*. In particular, both sums have all diagonal entries equal to* τ_T*.*

(c) $\sum_j \mathfrak{n}_j^{-1}\kappa_j B_j\mathbf{e} = \sum_j \mathfrak{n}_j^{-1}\kappa_j B_j^\circ\mathbf{e} = \tau_T\mathbf{e}$*.*

(d) *Let* S *be the permutation matrix that exchanges the coordinates x and y, and write $B_j(k)$ for the dependence on k. Then $\mathsf{S}B_j(k)\mathsf{S} = B_j(-k)$ and $\mathsf{S}B_j^\circ\mathsf{S} = B_j^\circ$ for every j.*

Proof. For (a), we have $\mathbf{e}\cdot\mathbf{c}_0 = 3$, $\mathbf{e}\cdot\mathbf{c}_1 = 2$, $\mathbf{e}\cdot\mathbf{c}_v = \mathbf{e}\cdot\mathbf{c}_x = \mathbf{e}\cdot\mathbf{c}_y = 1$ and $L_2\mathbf{e} = 0$, and the same numbers for the baseline vectors $\mathbf{e}$, $\mathbf{e}_x+\mathbf{e}_y$ and $\mathbf{e}_v$. For (b), note that $h\sum_{j\le4}\kappa_j B_j$ is the covariance of the perturbed window fields divided by b. The computation above gives

$$h\sum_{j\le4}\kappa_j B_j = h\sum_{j\le4}\kappa_j B_j^\circ + \delta\big(\mathbf{e}_v(\mathbf{e}_x-\mathbf{e}_y)^{\mathsf{T}} + (\mathbf{e}_x-\mathbf{e}_y)\mathbf{e}_v^{\mathsf{T}}\big) + hk^2 L_\vartheta L_2,$$

while $h\kappa_5 B_5 = h\kappa_5 B_5^\circ - hk^2 L_\vartheta L_2$. Adding and dividing by h gives (b). The diagonal of the baseline sum is $\kappa_1+\kappa_2+\kappa_4+\kappa_5 = \tau_T$ in the coordinates x, y and $\kappa_1+\kappa_3+\kappa_4+\kappa_5 = \tau_T$

in the coordinate v. For (c), we have $B_1\mathbf{e} = 3\mathbf{c}_0$, $B_2\mathbf{e} = 2\mathbf{c}_1$, $B_3\mathbf{e} = \mathbf{c}_v$, $B_4\mathbf{e} = \mathbf{c}_v + \mathbf{c}_x + \mathbf{c}_y$ and $B_5\mathbf{e} = \mathbf{e}$, so that

$$\sum_j \mathfrak{n}_j^{-1}\kappa_j B_j\mathbf{e} = \mathbf{c}_0 + \vartheta\mathbf{c}_1 + \mathbf{c}_v + (1-\vartheta)(\mathbf{c}_x + \mathbf{c}_y) + \kappa_5\mathbf{e}.$$

Its v entry is $2 + \kappa_5$. Its x entry is

$$\begin{aligned}(1+\omega_0 k) + \vartheta(1-\omega_0 k) + \omega_1 k + (1-\vartheta)(1-2\omega_0 k) + \kappa_5 \\ = 2 + \kappa_5 + k\big(\omega_1 - (1-\vartheta)\omega_0\big) = \tau_T,\end{aligned}$$

and the y entry is the same with k replaced by $-k$. For the baseline we find $\mathbf{e} + \vartheta(\mathbf{e}_x + \mathbf{e}_y) + \vartheta\mathbf{e}_v + (1-\vartheta)\mathbf{e} + \kappa_5\mathbf{e} = \tau_T\mathbf{e}$. Finally, (d) holds because $\mathsf{S}\mathbf{c}_0(k) = \mathbf{c}_0(-k)$, $\mathsf{S}\mathbf{c}_1(k) = \mathbf{c}_1(-k)$, $\mathsf{S}\mathbf{c}_v(k) = \mathbf{c}_v(-k)$, $\mathsf{S}\mathbf{c}_x(k) = \mathbf{c}_y(-k)$, $\mathsf{S}\mathbf{c}_y(k) = \mathbf{c}_x(-k)$ and $\mathsf{S}L_2\mathsf{S} = L_2$. □

Part (b) says that the window shifts the overlaps of v with x and y by $\pm\delta$ and changes nothing else in the total covariance. Part (c) is the identity behind the cancellation of the linear part of the terminal function in Sections 7.3 and 7.4.

Lemma 7.2 (Admissibility and endpoint)**.** *Let $q_0 \in K$, $0 \le h \le h_1$ and $|k| \le k_0$, and suppose that $\vartheta K_\beta h \le \alpha_*$. Then every increment of the perturbed path is positive semidefinite, and its masses are nondecreasing along the path. Its endpoint is*

$$D_\delta = D(r,q,\delta) = \begin{pmatrix} 1 & r+\delta & r-\delta \\ r+\delta & 1 & q \\ r-\delta & q & 1 \end{pmatrix}. \tag{7.6}$$

The same holds for the baseline path, whose endpoint is $D(r,q,0)$.

Proof. The rates $B_1, \dots, B_4$ and the baseline rates are sums of matrices of the form $\mathbf{c}\mathbf{c}^{\mathsf{T}}$. The matrix L_2 has the eigenvalue $|\mathbf{e}_x - \mathbf{e}_y|^2 = 2$ on $\mathbf{e}_x - \mathbf{e}_y$ and the eigenvalue 0 on the orthogonal complement, and $\mathrm{Id} - L_2/4$ has eigenvalues $1, 1, 1/2$. This proves that the increments are positive semidefinite.

The masses are nondecreasing within each stage because α is nondecreasing. At the junctions, they are $\alpha(q_0)/3$ on both sides of q_0, then $\alpha(r)/3 \le \alpha(r)/2$, then $\alpha(q)/2 \le \alpha(r)$, and they are continuous at the remaining junctions. The inequality $\alpha(q) \le 2\alpha(r)$ holds because $[q_0, T] \subset K'$, so that

$$\alpha(q) - \alpha(r) \le K_\beta(q-r) = \vartheta K_\beta h \le \alpha_* \le \alpha(r).$$

All masses are at most 1.

For the endpoint, the baseline path ends at $q_0\mathbf{e}\mathbf{e}^{\mathsf{T}} + h\sum_j \kappa_j B_j^\circ + (1-T)\mathrm{Id}$. We have $h\kappa_1 = h$, $h\kappa_2 = h\kappa_3 = \vartheta h$ and $h(\kappa_4 + \kappa_5) + 1 - T = 1 - q$, and this endpoint equals

$$r\,\mathbf{e}\mathbf{e}^{\mathsf{T}} + \vartheta h\big((\mathbf{e}_x + \mathbf{e}_y)(\mathbf{e}_x + \mathbf{e}_y)^{\mathsf{T}} + \mathbf{e}_v\mathbf{e}_v^{\mathsf{T}}\big) + (1-q)\mathrm{Id} = D(r,q,0).$$

By Lemma 7.1(b), the endpoint of the perturbed path is $D(r,q,0) + hk\big(\mathbf{e}_v(\mathbf{e}_x - \mathbf{e}_y)^{\mathsf{T}} + (\mathbf{e}_x - \mathbf{e}_y)\mathbf{e}_v^{\mathsf{T}}\big) = D_\delta$. □

The trial value

Let $\Lambda_* = \max_{|k|\le k_0} \sum_j \kappa_j |B_j|$, where $|B_j|$ is the operator norm, and fix $h_0 \in (0, h_1]$ such that

$$\vartheta K_\beta h_0 \le \alpha_*, \qquad 6bh_0\Lambda_* \le 1. \tag{7.7}$$

In the rest of this section we assume that $q_0 \in K$, $0 \le h \le h_0$ and $|k| \le k_0$, so that Lemma 7.2 applies. We now define the trial value and express it through the local window.

For $N \ge 1$, let $(Q^{(N)}, m^{(N)})$ be the admissible path obtained from the perturbed path by dividing $[0, q_0]$, each scalar interval I_j and $[T, 1]$ into N intervals of equal length, keeping the covariance increments, and giving each increment the mass of the path at the left endpoint of its scalar interval. The proof of Lemma 7.2 shows that these masses are nondecreasing, and the endpoint is D_δ. We define

$$\mathcal{T}(q_0, h, k) = \lim_{N\to\infty} \Big(\Psi\big(Q^{(N)}, m^{(N)}, 0\big) - \frac{b}{4} \int m^{(N)} \, \mathrm{d}\|Q^{(N)}\|_F^2 \Big). \tag{7.8}$$

The expression in parentheses is the trial value $\mathcal{T}_{D_\delta}(Q^{(N)}, m^{(N)}, 0)$ of Definition 4.15, whose mismatch term $\|D_\delta - Q_N\|_F^2$ vanishes because the endpoint is D_δ; by Lemma 7.3 below, $\mathcal{T}(q_0, h, k)$ is a trial value for D_δ in the sense of that definition. We write $\mathcal{T}^\circ(q_0, h, k)$ for the same limit for the baseline path.

To describe the limit, we use the flows of Section 6. The *local block* is the flow, in the sense of Definition 6.1 and Lemma 6.2, of the stage datum with $\ell = 3$, $N = 5$, matrices $B_1, \dots, B_5$, durations $h\kappa_1, \dots, h\kappa_5$, and masses $m_j(\tau) = \mathfrak{n}_j^{-1} \alpha\big(q_0 + h(o_j + \kappa_j \tau)\big)$ for $\tau \in [0, 1]$. Here $\tau = 0$ is the end of I_j closest to the root of the hierarchy, and $\tau = 1$ the end closest to the terminal time. The flow F solves

$$-\partial_\tau F = \frac{bh\kappa_j}{2} \Big\{ B_j : D^2 F + m_j(\tau - j + 1)\, \nabla F \cdot B_j \nabla F \Big\} \quad \text{on } (j-1, j) \times \mathbb{R}^3, \qquad j = 1, \dots, 5, \tag{7.9}$$

with $F(5, \mathbf{z}) = \sum_{i \in \{v,x,y\}} \Phi(T, z_i)$, and we set

$$\mathcal{B}(q_0, h, k, \mathbf{z}) = F(0, \mathbf{z}).$$

We define $\mathcal{B}^\circ$ in the same way with the rates B_j°. For masses that are constant on the pieces of a partition, the flow is the composition of the integration steps of Definition 4.1 (see (6.2)), that is, the value of a finite hierarchy; for the masses above it is the limit of these values (Lemma 6.2(b)).

For the initial segment, let $f : \mathbb{R} \to \mathbb{R}$ have bounded derivatives of orders one and two. We write $\mathcal{H}(f) = \mathcal{H}^\alpha_{q_0,3} f$, in the notation of Corollary 6.14: this is the value at $(0, 0)$ of the solution Φ_f^α of

$$-\partial_t \Phi_f^\alpha = \frac{b}{2} \Big(\partial_{zz} \Phi_f^\alpha + \frac{\alpha(t)}{3} (\partial_z \Phi_f^\alpha)^2 \Big) \quad \text{on } [0, q_0] \times \mathbb{R}, \qquad \Phi_f^\alpha(q_0, \cdot) = f,$$

in the sense of Lemma 6.2. It is the one-site recursion of three coordinates that share every field increment on $[0, q_0]$, acting on a function of their common value (Remark 6.15), and it is the operator $\mathcal{H}_\alpha$ of Section 8.4 on the interval $[0, q_0]$. By the first claim of Corollary 6.14,

$$\mathcal{H}\big(3\Phi(q_0, \cdot)\big) = 3\Phi(0, 0). \tag{7.10}$$

Finally, the penalties of the two paths are the limits $\Pi(q_0, h, k)$ and Π° of $\int m^{(N)}\, \mathrm{d}\|Q^{(N)}\|_F^2$, and we write $\Delta\Pi = \Pi - \Pi^\circ$ and

$$V(q_0, h, k, z) = \mathcal{B}(q_0, h, k, z\mathbf{e}) - 3\Phi(q_0, z) \qquad (z \in \mathbb{R}). \tag{7.11}$$

Lemma 7.3 (The trial value)**.** *The limit* (7.8) *exists, and the following hold.*

(a) *For every $n \geq 2$ such that D_δ is feasible, we have $p_{3,n}(D_\delta) \leq \mathcal{T}(q_0, h, k)$.*

(b) *$\mathcal{T}(q_0, h, k) = \mathcal{H}\big(\mathcal{B}(q_0, h, k, \cdot\, \mathbf{e})\big) - \frac{b}{4}\Pi(q_0, h, k)$, and the same holds for the baseline path with $\mathcal{B}^\circ$ and Π°.*

(c) *$\mathcal{B}^\circ(q_0, h, k, z\mathbf{e}) = 3\Phi(q_0, z)$ for all $z \in \mathbb{R}$, $\Pi^\circ = 6\int_0^1 t\alpha(t)\, \mathrm{d}t$, and $\mathcal{T}^\circ(q_0, h, k) = 3P_\beta$.*

(d) *We have*

$$\mathcal{T}(q_0, h, k) - 3P_\beta = \mathcal{H}\big(3\Phi(q_0, \cdot) + V\big) - \mathcal{H}\big(3\Phi(q_0, \cdot)\big) - \frac{b}{4}\Delta\Pi. \tag{7.12}$$

(e) *Let $\widetilde{Q}_0 = 0$ and $\widetilde{Q}_j = \sum_{i \leq j} \kappa_i B_i$, and define $\widetilde{Q}_j^\circ$ in the same way with the rates B_i°. Then*

$$\Delta\Pi = \sum_{j=1}^{5} \frac{2h^2\kappa_j}{\mathfrak{n}_j} \int_0^1 \alpha\big(q_0 + h(o_j + \kappa_j\tau)\big)\Big[(\widetilde{Q}_{j-1} + \kappa_j\tau B_j) : B_j - (\widetilde{Q}_{j-1}^\circ + \kappa_j\tau B_j^\circ) : B_j^\circ\Big]\, \mathrm{d}\tau. \tag{7.13}$$

Proof. We first identify the values of the finite paths with step flows, then pass to the limit, and then treat the baseline and the penalties.

Step 1: the finite paths as step flows. We consider the stage datum with seven stages: the initial segment, with matrix $\mathbf{e}\mathbf{e}^\mathsf{T}$, duration q_0 and mass $\alpha(q_0\tau)/3$; the five stages of the window; and the final segment, with matrix Id, duration $1 - T$ and mass $\alpha(T + (1 - T)\tau)$. Its terminal function is $\mathbf{z} \mapsto \sum_i \log(2\cosh z_i)$. The masses of $(Q^{(N)}, m^{(N)})$ are the sampled masses of this datum for the partition of each stage into N pieces of equal length. By the remark after (6.2), $\Psi(Q^{(N)}, m^{(N)}, 0)$ is the output at $\mathbf{z} = 0$ of the corresponding step flow.

Step 2: the limit. The sampled masses converge in L^1 to the masses of the datum, because these are nondecreasing functions. By Lemma 6.2(b), the step flows converge uniformly to the flow of the seven-stage datum, and in particular the values converge. By Lemma 6.2(c), applied twice, this flow is obtained by composing the flows of the three parts. The final segment, with this terminal function, gives $\sum_i \Phi(T, z_i)$: its matrix is Id

and the terminal function is a sum of functions of single coordinates; by Lemma 4.2(c), its step flows are sums of scalar recursions (4.3), which converge to Φ by the definition of Φ_α (Proposition 4.3(i)). This function belongs to $\mathcal{C}^2$, since $|u| \le 1$ and $0 < \chi \le 1$, and to $\mathcal{C}^\infty$ by Lemma 5.1. The window then gives $\mathcal{B}$, which belongs to $\mathcal{C}^2$ by Lemma 6.2(d): its terminal function has $\|D^2 f\|_\infty \le 1$, and $2 \cdot 3\, b\, \|D^2 f\|_\infty \sum_j h\kappa_j |B_j| \le 6bh_0\Lambda_* \le 1$ by (7.7). By Remark 6.15 the initial segment acts on $z \mapsto \mathcal{B}(z\mathbf{e})$ as $\mathcal{H}$. We conclude that the values converge to $\mathcal{H}(\mathcal{B}(\cdot\,\mathbf{e}))$. The penalties converge as well. On the window, this is Lemma 6.7, with $Q_\star = q_0\mathbf{e}\mathbf{e}^{\mathsf{T}}$. On the initial and final segments, the penalties are Riemann–Stieltjes sums $\sum_i m_i(\|Q_i\|_F^2 - \|Q_{i-1}\|_F^2)$, in which m_i is the value of a nondecreasing function at the left endpoint of a piece and $\|Q\|_F^2$ is a continuous, piecewise quadratic function of the scalar time; they converge to the corresponding integrals. This shows that the limit (7.8) exists, and it proves (b). The endpoints, values and penalties converge, and (a) follows from the last sentence of Proposition 4.19.

Step 3: the baseline. Let α_N be the step distribution function equal to α at the left endpoints of the scalar partition. The baseline finite path is the path of Lemma 4.20(b) for α_N, whose breakpoints contain r and q: the common increments are those on $[0, r]$, the xy-increments of stage 2 precede the v-increments of stage 3 (their masses are ordered by Lemma 7.2), and the independent increments are those on $[q, 1]$. Lemma 4.20(b) implies that its trial value is $3\mathcal{P}_\beta(\alpha_N)$ and that its penalty is $6\int_0^1 t\alpha_N(t)\,\mathrm{d}t$. By Proposition 4.3(ii), these converge to $3\mathcal{P}_\beta(\alpha)$ and $6\int_0^1 t\alpha(t)\,\mathrm{d}t$, and $\mathcal{P}_\beta(\alpha) = P_\beta$ by Theorems 4.4 and 4.5. The proof of Lemma 4.20(b) also shows that the step flow of the baseline window, applied to $\sum_i \Phi_{\alpha_N}(T, z_i)$, equals $3\Phi_{\alpha_N}(q_0, z)$ at $z\mathbf{e}$. By the last estimate of Lemma 6.2(a), by Lemma 6.2(b), and by Proposition 4.3(ii), the left side converges to $\mathcal{B}^\circ(q_0, h, k, z\mathbf{e})$ and the right side to $3\Phi(q_0, z)$. This proves (c), and (d) follows from (b), (c) and (7.10).

Step 4: the penalty. During stage j, at the fraction $\tau \in [0, 1]$ of its scalar interval, the perturbed path is $Q = q_0\mathbf{e}\mathbf{e}^{\mathsf{T}} + h(\widetilde{Q}_{j-1} + \kappa_j\tau B_j)$. Then $\mathrm{d}\|Q\|_F^2 = 2Q : \mathrm{d}Q = 2h\kappa_j Q : B_j\,\mathrm{d}\tau$, and the mass is $\alpha(q_0 + h(o_j + \kappa_j\tau))/\mathfrak{n}_j$. The term $2h\kappa_j q_0 \mathbf{e}\mathbf{e}^{\mathsf{T}} : B_j = 2h\kappa_j q_0\, \mathbf{e} \cdot B_j\mathbf{e}$ is the same for the baseline path, by Lemma 7.1(a). The initial segments of the two paths coincide. On $[T, 1]$, both paths have $\mathrm{d}\|Q\|_F^2 = 2\,\mathrm{tr}\,Q\,\mathrm{d}t$, and $\mathrm{tr}\,Q(T)$ is the same for both paths, because $\sum_j \kappa_j B_j$ and $\sum_j \kappa_j B_j^\circ$ have the same diagonal (Lemma 7.1(b)). Subtracting the two penalties, we obtain (7.13). □

By Lemma 7.3(a), every upper bound for $\mathcal{T}(q_0, h, k)$ is an upper bound for the constrained pressure $p_{3,n}(D_\delta)$. The rest of the section bounds $\mathcal{T}(q_0, h, k) - 3P_\beta$ through (7.12). In (7.12), the function V, the number $\Delta\Pi$, and the part of Π accumulated on $[q_0, 1]$ depend on α only through its restriction to $[q_0, 1]$. Changing α on $[0, q_0)$ leaves them unchanged; we use this in Section 8.

7.2 Derivatives of the local block

We now apply the estimates of Section 6 to the local block. Let $\mathcal{U} = [0, h_1] \times [-k_0, k_0]$. The stage datum of (7.9) satisfies the hypotheses (H1)–(H5) of Section 6, with $\ell = 3$ and $N = 5$, for the following reasons.

- (H1): the matrices B_j and B_j° do not depend on h, are polynomial in k, and are positive semidefinite (Lemma 7.2). Some of them are singular.
- (H2): the durations are $h\kappa_j(k)$, where $\kappa_j \geq 0$ is a polynomial in k. The coefficient $\kappa_5 = 4k^2 L_\vartheta$ of the terminal correction vanishes at $k = 0$.
- (H3): the masses $m_j(\tau, h, k) = \mathfrak{n}_j^{-1} \alpha(q_0 + h(o_j + \kappa_j(k)\tau))$ take values in $[0, 1]$. For $(h, k) \in \mathcal{U}$, the argument of α lies in K', on which α is infinitely differentiable, and all derivatives of m_j in (h, k) exist and are continuous.
- (H4): the terminal function is $f(h, k, \mathbf{z}) = \sum_i \Phi(q_0 + h\tau_T(k), z_i)$. By Lemma 5.1 and Lemma 5.3(v), every derivative $\partial_h^{\gamma_h} \partial_k^{\gamma_k} D_{\mathbf{z}}^\iota f$ with $(\gamma_h, \gamma_k, \iota) \neq (0, 0, 0)$ exists, is continuous, and is bounded on $\mathcal{U} \times \mathbb{R}^3$.
- (H5): since $0 < \chi \leq 1$ and f is a sum of functions of single coordinates, we have $\|D^2 f\|_\infty \leq 1$, and $2\ell b h_0 \Lambda_* \|D^2 f\|_\infty \leq 6 b h_0 \Lambda_* \leq 1$ by (7.7).

The suprema that enter the constants of Proposition 6.3(ii) and Lemma 6.7 are bounded uniformly in $q_0 \in K$, because they involve only the derivatives of α and Φ on K'. By Remark 6.6, the conclusions are uniform in $q_0 \in K$.

Lemma 7.4 (Derivatives of the local block)**.** *Let h_0 be as in* (7.7)*. The following hold for $q_0 \in K$, $0 \leq h \leq h_0$ and $|k| \leq k_0$, where derivatives in h at $h = 0$ are right derivatives. All constants depend only on β, ϑ, k_0, K and the orders of the derivatives.*

(i) *For all integers $\gamma_h, \gamma_k, j \geq 0$ with $\gamma_h + \gamma_k + j \geq 1$, the derivatives $\partial_h^{\gamma_h} \partial_k^{\gamma_k} D_{\mathbf{z}}^j \mathcal{B}$ and $\partial_h^{\gamma_h} \partial_k^{\gamma_k} D_{\mathbf{z}}^j \mathcal{B}^\circ$ exist, are continuous, and are bounded uniformly in $\mathbf{z} \in \mathbb{R}^3$.*

(ii) *For all integers $\gamma_h, \gamma_k, j \geq 0$, the derivative $\partial_h^{\gamma_h} \partial_k^{\gamma_k} \partial_z^j V(q_0, h, k, z)$ is bounded uniformly in z. Moreover,*

$$V(q_0, h, 0, \cdot) = 0, \qquad V(q_0, h, -k, \cdot) = V(q_0, h, k, \cdot),$$

and $V(q_0, 0, k, \cdot) = \partial_h V(q_0, 0, k, \cdot) = 0$.

(iii) *There exists C such that, for $0 \leq j \leq 3$ and all z,*

$$|\partial_z^j V(q_0, h, k, z)| \leq C h^2 k^2. \tag{7.14}$$

(iv) *The function $\Delta\Pi$ is smooth in (h, k), with bounded derivatives of every order, it is even in k, it vanishes at $k = 0$, and $|\Delta\Pi| \leq C h^2 k^2$.*

We have $\|V\|_{C_b^3} \leq C h^2 k^2 = C\delta^2$, which is at most $C\delta^2/h$ for $h \leq 1$. Of the derivatives in (i) and (ii), we use those with $\gamma_h \leq 4$, $\gamma_k \leq 2$, and $j \leq 3$; they have weight $j + 2(\gamma_h + \gamma_k) \leq 15$ in the sense of Section 6.

Proof. Part (i) is Proposition 6.3(ii), with constants uniform in $q_0 \in K$ by Remark 6.6. The bounds include the right derivatives at $h = 0$ and the derivatives at $k = 0$, where the terminal correction has zero duration (Remark 6.5). The value $\mathcal{B}$ itself has linear growth, and (i) does not bound it.

For (ii), we first prove the identities. If $k = 0$, then $B_j = B_j^\circ$ for $j \le 4$ and $\kappa_5 = 0$. The perturbed and baseline windows coincide, and $V = \mathcal{B}^\circ(\cdot\,\mathbf{e}) - 3\Phi(q_0, \cdot) = 0$ by Lemma 7.3(c). For the evenness, let S be as in Lemma 7.1(d). The durations and masses are even in k, and the terminal function is invariant under S. If F solves (7.9) at k, then $(\tau, \mathbf{z}) \mapsto F(\tau, \mathsf{S}\mathbf{z})$ solves it at $-k$, by Lemma 7.1(d), and it has bounded first and second spatial derivatives. By the uniqueness in Proposition 6.3(i), $\mathcal{B}(q_0, h, -k, \mathbf{z}) = \mathcal{B}(q_0, h, k, \mathsf{S}\mathbf{z})$, and $\mathsf{S}(z\mathbf{e}) = z\mathbf{e}$.

At $h = 0$ all durations vanish. Let $\varphi(\mathbf{z}) = \sum_i \Phi(q_0, z_i)$. Then $\mathcal{B}(q_0, 0, k, \mathbf{z}) = \varphi(\mathbf{z})$ by Remark 6.4, which implies that $V(q_0, 0, k, \cdot) = 0$. By (6.5), in which the average mass of stage j at $h = 0$ is $a/\mathfrak{n}_j$ with $a = \alpha(q_0)$,

$$\partial_h \mathcal{B}(q_0, 0, k, \mathbf{z}) = \tau_T \sum_i \partial_t \Phi(q_0, z_i) + \frac{b}{2} \sum_{j=1}^{5} \kappa_j \Big\{ B_j : D^2\varphi(\mathbf{z}) + \frac{a}{\mathfrak{n}_j} \nabla\varphi(\mathbf{z}) \cdot B_j \nabla\varphi(\mathbf{z}) \Big\}.$$

The first term comes from the dependence of T on h. At $\mathbf{z} = z\mathbf{e}$, we have $D^2\varphi = \chi(q_0, z)\mathrm{Id}$ and $\nabla\varphi = u(q_0, z)\mathbf{e}$, and the sum equals $\chi(q_0, z) \sum_j \kappa_j \operatorname{tr} B_j + u(q_0, z)^2 a \sum_j \mathfrak{n}_j^{-1} \kappa_j\, \mathbf{e} \cdot B_j \mathbf{e}$. By Lemma 7.1(a) and (b), it does not change when B_j is replaced by B_j°. Since $\mathcal{B}^\circ(q_0, h, k, z\mathbf{e}) = 3\Phi(q_0, z)$ does not depend on h, we conclude that $\partial_h V(q_0, 0, k, \cdot) = 0$. The bounds in (ii) with $\gamma_h + \gamma_k + j \ge 1$ follow from (i), the formula for derivatives along the diagonal in Remark 6.6, and Lemma 5.1, since V is the difference of $\mathcal{B}(\cdot\,\mathbf{e})$ and the function $3\Phi(q_0, \cdot)$, which does not depend on (h, k). Finally, V itself is bounded, since $V(q_0, 0, k, \cdot) = 0$ and $\partial_h V$ is bounded.

For (iii), let $0 \le j \le 3$. Taylor's formula and the identities $V(q_0, 0, k, \cdot) = \partial_h V(q_0, 0, k, \cdot) = 0$ yield $\partial_z^j V(h) = \int_0^h (h - h')\, \partial_h^2 \partial_z^j V(h')\, \mathrm{d}h'$. The function $\partial_h^2 \partial_z^j V(h', k)$ vanishes at $k = 0$ and is even in k, which implies that $|\partial_h^2 \partial_z^j V(h', k)| \le \frac{k^2}{2} \sup |\partial_k^2 \partial_h^2 \partial_z^j V|$. Combining these bounds with (ii), we obtain (7.14).

For (iv), Lemma 6.7 with $Q_\star = q_0 \mathbf{e}\mathbf{e}^\mathsf{T}$ shows that the penalties of the two windows are smooth with bounded derivatives of every order, and their contributions on $[0, q_0]$ and $[T, 1]$ cancel in $\Delta\Pi$ (Step 4 of the proof of Lemma 7.3). The penalty of each path is invariant under $Q \mapsto \mathsf{S}Q\mathsf{S}$, since $\|\mathsf{S}Q\mathsf{S}\|_F = \|Q\|_F$; by Lemma 7.1(d), this proves that $\Delta\Pi$ is even in k. At $k = 0$ the two windows coincide, and $\Delta\Pi = 0$. Finally, $\Delta\Pi = h^2 \Theta(h, k)$, where Θ is the sum in (7.13) without the factor h^2; since Θ is smooth, even in k and zero at $k = 0$, we have $|\Theta| \le Ck^2$. □

Corollary 7.5 (Reduction to the boundary perturbation). *For $q_0 \in K$, $0 \le h \le h_0$, and $|k| \le k_0$,*

$$\mathcal{T}(q_0,h,k) - 3P_\beta = \mathbb{E}V(q_0,h,k,Y_{q_0}) - \frac{b}{4}\Delta\Pi(q_0,h,k) + \mathcal{E}_{\mathrm{in}}, \qquad |\mathcal{E}_{\mathrm{in}}| \le Ch^4k^4.$$

Proof. By Lemma 7.4(ii), V, V', and V'' are bounded, and $\|V\|_\infty \le Ch^2k^2$ by (7.14). Using Corollary 6.14 with $\mathfrak{n} = 3$, $t_\star = q_0$, and $\varepsilon = Ch^2k^2$, we find that

$$\Big|\mathcal{H}\big(3\Phi(q_0,\cdot) + V\big) - 3\Phi(0,0) - \mathbb{E}V(q_0,h,k,Y_{q_0})\Big| \le \frac{C^2h^4k^4}{2}.$$

We have $3\Phi(0,0) = \mathcal{H}(3\Phi(q_0,\cdot))$ by (7.10), and inserting this bound into (7.12) completes the proof. □

The expectation $\mathbb{E}V(q_0,h,k,Y_{q_0})$ is taken under the law of the scalar diffusion at the fixed time q_0, which does not depend on h or k. This is why we expand at a fixed left endpoint q_0 rather than at a fixed center r.

We expand $\Delta\Pi$ explicitly. We write $a = \alpha(q_0)$ and insert $\alpha(q_0 + h(o_j + \kappa_j\tau)) = a + \alpha'(q_0)h(o_j + \kappa_j\tau) + O(h^2)$ in (7.13). Then the Taylor polynomial of degree three of $\Delta\Pi$ in h is $ah^2N_\vartheta(k) + \alpha'(q_0)h^3N_\vartheta^{(1)}(k)$, where

$$N_\vartheta^{(i)}(k) = \sum_{j=1}^{5} \frac{2\kappa_j}{\mathfrak{n}_j} \int_0^1 (o_j + \kappa_j\tau)^i \Big[(\widetilde{Q}_{j-1} + \kappa_j\tau B_j) : B_j - (\widetilde{Q}_{j-1}^{\circ} + \kappa_j\tau B_j^{\circ}) : B_j^{\circ}\Big]\,\mathrm{d}\tau \tag{7.15}$$

for $i \in \{0,1\}$, and $N_\vartheta = N_\vartheta^{(0)}$. The integrands are polynomials in τ, and we find

$$N_\vartheta(k) = \frac{8(1-\vartheta)}{(2-\vartheta)^2}k^2 - \frac{2(3\vartheta^2+6\vartheta+4)}{3(2-\vartheta)^4}k^4, \qquad N_{1/2}(k) = \frac{16}{9}k^2 - \frac{248}{243}k^4, \tag{7.16}$$

and $N_{1/2}^{(1)}(k) = \frac{103}{108}k^2 - \frac{7180}{729}k^4 - \frac{21296}{729}k^6$. The polynomial $N_\vartheta^{(1)}$ enters only through the cancellation in Proposition 7.13, and we do not use its value.

7.3 Taylor coefficients at constant mass

By Lemma 7.4, the function $h \mapsto V(q_0,h,k,z)$ has four bounded derivatives, and it is determined up to an error $O(h^4)$ by its Taylor polynomial of degree three. In this section and the next we compute this polynomial. We write $[X]_n = \partial_h^n X|_{h=0}/n!$ for the Taylor coefficients of a function X of h at $h = 0$.

The recursion for the coefficients

By Proposition 6.3(iii), the derivatives $\partial_h^n F$ solve the equations obtained by differentiating (7.9) in h. Evaluating them at $h = 0$, we obtain a recursion for the Taylor coefficients $[F]_n(\tau, \mathbf{z})$ of the flow. We write $[m_j]_l(\tau) = \alpha^{(l)}(q_0)\,(o_j + \kappa_j\tau)^l/(l!\,\mathfrak{n}_j)$ for the Taylor coefficients of the masses. The right side of (7.9) contains the factor h, and taking the coefficient of h^n we find, for $n \geq 1$ and $\tau \in [j-1, j]$,

$$[F]_n(\tau) = [F]_n(j) + \frac{b\kappa_j}{2}\int_\tau^j \Big\{ B_j : D^2[F]_{n-1}(\tau') + \sum_{l+p+p'=n-1} [m_j]_l(\tau' - j + 1)\,\nabla[F]_p(\tau')\cdot B_j \nabla[F]_{p'}(\tau') \Big\}\,\mathrm{d}\tau', \tag{7.17}$$

with $[F]_0 = f(0, k, \cdot)$ and $[F]_n(5, \mathbf{z}) = \frac{\tau_T^n}{n!}\sum_i \partial_t^n \Phi(q_0, z_i)$. We call each use of the right side of (7.9) in this recursion an *insertion*. By (5.19) and its consequences in the proof of Lemma 5.3, $\partial_t^n\Phi(q_0, \cdot)$ is a polynomial in $\alpha(q_0), \dots, \alpha^{(n-1)}(q_0)$ and in the spatial derivatives of $\Phi(q_0, \cdot)$ of orders one to $2n$. By induction on n, for every $n \geq 1$ the coefficient $[\mathcal{B}]_n(\mathbf{z})$ is a sum of products of spatial derivatives of positive order of $\Phi(q_0, \cdot)$ at z_v, z_x, z_y, of total spatial order at most $2n$, with coefficients that are polynomials in k and in $\alpha(q_0), \dots, \alpha^{(n-1)}(q_0)$. Each power h^l of the expansion of a mass carries a factor $\alpha^{(l)}(q_0)$, and each insertion of the right side of (7.9) carries a factor h; in particular, $\alpha'''(q_0)$ does not occur for $n \leq 3$. The same holds for $\mathcal{B}^\circ$ and for V.

We call the *constant-mass part* of $[V]_n$ the value of this polynomial at $\alpha'(q_0) = \alpha''(q_0) = 0$, with the other variables unchanged. It is the value of the recursion when every mass m_j is replaced by the constant $a/\mathfrak{n}_j$, with $a = \alpha(q_0)$, and when the time derivatives of Φ at q_0 are computed from the Parisi equation with the constant profile a. In this section we compute the constant-mass part. The remaining part is computed in Section 7.4.

Formal Gaussian steps

With constant masses, each stage is a Gaussian integration step, and the three stages of mass a combine into one step by Lemma 4.2(b). It is convenient to subtract the scalar evolution of the terminal function from this step. The result is a step whose covariance is not positive semidefinite, and we treat it as a formal object.

Let $\ell \geq 1$, let Σ be a symmetric $\ell \times \ell$ matrix, let $m \in [0, 1]$, and let f be a smooth function on $\mathbb{R}^\ell$. We define $f^{[0]} = f$ and, for $n \geq 1$,

$$f^{[n]} = \frac{1}{2n}\Big(\Sigma : D^2 f^{[n-1]} + m \sum_{p+p'=n-1} \nabla f^{[p]}\cdot\Sigma\nabla f^{[p']}\Big), \tag{7.18}$$

and we write $\widehat{T}_{m,\varepsilon\Sigma} f = \sum_{n\geq 0}\varepsilon^n f^{[n]}$ for the resulting formal power series in ε. It is the unique formal solution of $\partial_\varepsilon F = \frac{1}{2}\big(\Sigma : D^2 F + m\nabla F\cdot\Sigma\nabla F\big)$ with $F = f$ at $\varepsilon = 0$. By

induction, $f^{[n]}$ is a sum of products of derivatives of f of positive order, of total order $2n$, and it is a homogeneous polynomial of degree n in the entries of Σ. Formal series of functions are composed by substitution: if the function in (7.18) is itself a formal power series in ε, we insert this series and collect the powers of ε. The coefficient of each power involves finitely many terms. The relation with the integration steps of Definition 4.1 is the following.

Lemma 7.6 (Formal steps)**.** *Let $m \in [0,1]$, let Σ, Σ' be symmetric $\ell \times \ell$ matrices, and let f be smooth with bounded derivatives of positive order.*

(a) *If $\Sigma \succeq 0$, then $\varepsilon \mapsto T_{m,\varepsilon\Sigma} f(\mathbf{z})$ is infinitely differentiable on $[0,1]$, and $\widehat{T}_{m,\varepsilon\Sigma} f(\mathbf{z})$ is its Taylor series at $\varepsilon = 0$.*

(b) $\widehat{T}_{m,\varepsilon\Sigma} \widehat{T}_{m,\varepsilon\Sigma'} f = \widehat{T}_{m,\varepsilon(\Sigma+\Sigma')} f$.

(c) *For $c > 0$ such that $m/c \in [0,1]$, we have $\widehat{T}_{m/c,\varepsilon\Sigma}(cf) = c\, \widehat{T}_{m,\varepsilon\Sigma} f$.*

(d) *If Σ is diagonal with entries $\sigma_1, \dots, \sigma_\ell$ and $f(\mathbf{z}) = \sum_i \psi_i(z_i)$, then $\widehat{T}_{m,\varepsilon\Sigma} f(\mathbf{z}) = \sum_i \widehat{T}_{m,\varepsilon\sigma_i} \psi_i(z_i)$.*

(e) *For $c, u \in \mathbb{R}$ and $\mathbf{e}' \in \mathbb{R}^\ell$,*

$$\widehat{T}_{m,\varepsilon\Sigma}\big(c + u\,\mathbf{e}'\cdot\mathbf{z} + f\big)(\mathbf{z}) = c + u\,\mathbf{e}'\cdot\mathbf{z} + \frac{\varepsilon m u^2}{2}\,\mathbf{e}'\cdot\Sigma\mathbf{e}' + \big(\widehat{T}_{m,\varepsilon\Sigma} f\big)\big(\mathbf{z} + \varepsilon m u \Sigma \mathbf{e}'\big),$$

where the last term is expanded in powers of ε by Taylor's formula in the shift.

Proof. For (a), let $\Sigma \succeq 0$. By Lemma 6.9(b) and (c), the function $F(\varepsilon, \mathbf{z}) = T_{m,\varepsilon\Sigma} f(\mathbf{z})$ is infinitely differentiable on $[0,1] \times \mathbb{R}^\ell$ and solves $\partial_\varepsilon F = \frac{1}{2}(\Sigma : D^2 F + m \nabla F \cdot \Sigma \nabla F)$. Differentiating this equation $n-1$ times in ε and setting $\varepsilon = 0$, we find by induction that $\partial_\varepsilon^n F(0,\cdot) = n!\, f^{[n]}$.

For (b), both sides are formal series whose coefficients are polynomials in the entries of Σ and Σ' and in the derivatives of f. If $\Sigma, \Sigma' \succeq 0$, then Lemma 4.2(b), part (a), and the chain rule show that the two sides have the same coefficients. The set of pairs of positive semidefinite matrices contains a nonempty open subset of the space of pairs of symmetric matrices, and a polynomial that vanishes on a nonempty open set vanishes identically. This proves (b). Parts (c) and (d) follow from (7.18) by induction on n: for (c), the function $(cf)^{[n]}$ computed with mass m/c equals $c\, f^{[n]}$ computed with mass m, since the quadratic term acquires the factor $(m/c)c^2 = mc$; for (d), mixed derivatives of f vanish and the recursion splits into one recursion for each coordinate. For (e), let $\Sigma \succeq 0$ and $m > 0$, and let $\mathbf{X}$ be a centered Gaussian vector with covariance $\varepsilon\Sigma$. Multiplying the law of $\mathbf{X}$ by $e^{mu\mathbf{e}'\cdot\mathbf{X}} / \mathbb{E} e^{mu\mathbf{e}'\cdot\mathbf{X}}$ shifts its mean to $\varepsilon m u \Sigma \mathbf{e}'$, and $\mathbb{E} e^{mu\mathbf{e}'\cdot\mathbf{X}} = e^{\varepsilon m^2 u^2 \mathbf{e}'\cdot\Sigma\mathbf{e}'/2}$. We obtain

$$T_{m,\varepsilon\Sigma}\big(c + u\,\mathbf{e}'\cdot\mathbf{z} + f\big)(\mathbf{z}) = c + u\,\mathbf{e}'\cdot\mathbf{z} + \frac{\varepsilon m u^2}{2}\,\mathbf{e}'\cdot\Sigma\mathbf{e}' + \big(T_{m,\varepsilon\Sigma} f\big)\big(\mathbf{z} + \varepsilon m u \Sigma \mathbf{e}'\big). \tag{7.19}$$

For $m = 0$ this is immediate. Taking Taylor series in ε and arguing as in (b), we obtain (e) for every symmetric Σ. □

The identity (7.19) is the Cameron–Martin formula for a finite-dimensional Gaussian shift, obtained by completing the square as in [107, Section A.3, (A.6)].

The coefficient matrices

We define

$$\Lambda_0 = \mathbf{c}_0\mathbf{c}_0^{\mathsf{T}}, \qquad \Lambda_1 = \vartheta\,\mathbf{c}_1\mathbf{c}_1^{\mathsf{T}}, \qquad \Lambda_2 = \mathbf{c}_v\mathbf{c}_v^{\mathsf{T}} + (1-\vartheta)\big(\mathbf{c}_x\mathbf{c}_x^{\mathsf{T}} + \mathbf{c}_y\mathbf{c}_y^{\mathsf{T}}\big) - 2\mathrm{Id} - k^2 L_\vartheta L_2. \tag{7.20}$$

We have $\Lambda_0 = \kappa_1 B_1$, $\Lambda_1 = \kappa_2 B_2$, and $\Lambda_2 = \kappa_3 B_3 + \kappa_4 B_4 + \kappa_5 B_5 - \tau_T \mathrm{Id}$. The matrices Λ_0 and Λ_1 are positive semidefinite, and Λ_2 is not: its (v,v) entry is -1. At $k = 0$ we have $\Lambda_0 = \mathbf{e}\mathbf{e}^{\mathsf{T}}$, $\Lambda_1 = \vartheta(\mathbf{e}_x + \mathbf{e}_y)(\mathbf{e}_x + \mathbf{e}_y)^{\mathsf{T}}$ and $\Lambda_2 = \vartheta\mathbf{e}_v\mathbf{e}_v^{\mathsf{T}} - (1+\vartheta)\mathrm{Id}$, which are the same expressions for the baseline rates. By Lemma 7.1(b) and (c),

$$\operatorname{diag}(\Lambda_0 + \Lambda_1 + \Lambda_2) = 0, \qquad \big(\tfrac{1}{3}\Lambda_0 + \tfrac{1}{2}\Lambda_1 + \Lambda_2\big)\mathbf{e} = 0. \tag{7.21}$$

Moreover, $\mathbf{e}\cdot\Lambda_0\mathbf{e} = 9$, $\mathbf{e}\cdot\Lambda_1\mathbf{e} = 4\vartheta$, $\mathbf{e}\cdot\Lambda_2\mathbf{e} = -3 - 2\vartheta$, and

$$\Lambda_0 + \Lambda_1 + \Lambda_2 = \begin{pmatrix} 0 & 1+k & 1-k \\ 1+k & 0 & 1+\vartheta \\ 1-k & 1+\vartheta & 0 \end{pmatrix}.$$

For $\vartheta = 1/2$ we have $\omega_0 = 2/3$, $\omega_1 = 1/3$, $L_{1/2} = 11/9$, $\Lambda_0 = \mathbf{c}_0\mathbf{c}_0^{\mathsf{T}}$ with $\mathbf{c}_0 = (1, 1+\frac{2k}{3}, 1-\frac{2k}{3})^{\mathsf{T}}$, and

$$\Lambda_2 = \begin{pmatrix} -1 & \frac{k}{3} & -\frac{k}{3} \\ \frac{k}{3} & -\frac{2k^2}{3} - \frac{2k}{3} - \frac{3}{2} & \frac{2k^2}{3} \\ -\frac{k}{3} & \frac{2k^2}{3} & -\frac{2k^2}{3} + \frac{2k}{3} - \frac{3}{2} \end{pmatrix},$$

$$\Lambda_0 + \Lambda_1 = \begin{pmatrix} 1 & 1+\frac{2k}{3} & 1-\frac{2k}{3} \\ 1+\frac{2k}{3} & \frac{2k^2}{3} + \frac{2k}{3} + \frac{3}{2} & \frac{3}{2} - \frac{2k^2}{3} \\ 1-\frac{2k}{3} & \frac{3}{2} - \frac{2k^2}{3} & \frac{2k^2}{3} - \frac{2k}{3} + \frac{3}{2} \end{pmatrix}.$$

Lemma 7.7 (Reduction to three formal steps)**.** *Let $a = \alpha(q_0)$ and $\varphi(\mathbf{z}) = \sum_i \Phi(q_0, z_i)$. For $n \le 3$, the constant-mass part of $[\mathcal{B}]_n$ is the coefficient of h^n in*

$$\widehat{T}_{a/3,\,bh\Lambda_0}\,\widehat{T}_{a/2,\,bh\Lambda_1}\,\widehat{T}_{a,\,bh\Lambda_2}\,\varphi, \tag{7.22}$$

and the constant-mass part of $[\mathcal{B}^\circ]_n$ is the coefficient of h^n in the same expression at $k = 0$.

Proof. With the constant mass $a/\mathfrak{n}_j$, stage j maps $F(j,\cdot)$ to $F(j-1,\cdot) = T_{a/\mathfrak{n}_j, bh\kappa_j B_j} F(j,\cdot)$, by (6.2). By Lemma 7.6(a) and the chain rule, the recursion (7.17) for this stage is therefore the recursion of the formal step $\widehat{T}_{a/\mathfrak{n}_j, bh\kappa_j B_j}$, with $\varepsilon = h$. By Lemma 7.6(b), the stages 3, 4, and 5, all of mass a, combine into $\widehat{T}_{a,\, bh(\Lambda_2+\tau_T \mathrm{Id})}$. With the constant profile a, the Parisi equation reads $\partial_t \Phi = -\frac{b}{2}(\partial_{zz}\Phi + a(\partial_z \Phi)^2)$, and the Taylor series $\sum_n (h\tau_T)^n \partial_t^n \Phi(q_0,\cdot)/n!$ is the formal solution of this equation in the time $h\tau_T$, that is, $\widehat{T}_{a,-bh\tau_T}\Phi(q_0,\cdot)$. By Lemma 7.6(d), the terminal function is therefore $\widehat{T}_{a,-bh\tau_T \mathrm{Id}}\varphi$. By Lemma 7.6(b) again, $\widehat{T}_{a,bh(\Lambda_2+\tau_T\mathrm{Id})}\widehat{T}_{a,-bh\tau_T\mathrm{Id}} = \widehat{T}_{a,bh\Lambda_2}$, which proves the first claim. The second is the same computation with the baseline rates, whose coefficient matrices are the values of $\Lambda_0, \Lambda_1, \Lambda_2$ at $k = 0$. □

The scalar evolution of the terminal function between q_0 and T appears in (7.22) through the term $-\tau_T \mathrm{Id}$ in Λ_2. It accounts for the change of the curvature of the terminal function between the times T and q_0. This change is of order h, and it enters at order h^3 through the quadratic part of the terminal function; the term $-\tau_T \mathrm{Id}$ in Λ_2 retains this contribution in the coefficients computed below.

Expansion at a centered point

We now expand compositions of three formal steps applied to a separable function whose linear part vanishes. For symmetric matrices $\Sigma_0, \Sigma_1, \Sigma_2$ we write $\Sigma_{\le j} = \Sigma_0 + \cdots + \Sigma_j$ and $\Sigma_{\le -1} = 0$.

Lemma 7.8 (Expansion at a centered point)**.** *Let $\Sigma_0, \Sigma_1, \Sigma_2$ be symmetric 3×3 matrices with $\mathrm{diag}(\Sigma_{\le 2}) = 0$, and let $m_0, m_1, m_2 \in (0,1]$. Let $\psi\colon \mathbb{R}\to\mathbb{R}$ be smooth with $\psi(0) = \psi'(0) = 0$, write $\chi = \psi''(0)$ and $w = \psi'''(0)$, and set $\bar\psi(\mathbf{y}) = \sum_i \psi(y_i)$. Let $\mathfrak{E}(\varepsilon)$ be the value at $\mathbf{y} = 0$ of $\widehat{T}_{m_0,\varepsilon\Sigma_0}\widehat{T}_{m_1,\varepsilon\Sigma_1}\widehat{T}_{m_2,\varepsilon\Sigma_2}\bar\psi$. Then $[\mathfrak{E}]_0 = [\mathfrak{E}]_1 = 0$,*

$$[\mathfrak{E}]_2 = \frac{\chi^2}{4}\sum_{j=0}^{2} m_j\big(\|\Sigma_{\le j}\|_F^2 - \|\Sigma_{\le j-1}\|_F^2\big), \tag{7.23}$$

and

$$[\mathfrak{E}]_3 = \chi^3 \mathcal{Q} + \frac{w^2}{12}\sum_{j=0}^{2} m_j\big(\mathrm{cub}(\Sigma_{\le j}) - \mathrm{cub}(\Sigma_{\le j-1})\big), \tag{7.24}$$

where, with $\Upsilon_1 = m_2\Sigma_2 + m_1\Sigma_1$ and $\Upsilon_2 = m_2^2\Sigma_2^2 + m_1 m_2(\Sigma_1\Sigma_2 + \Sigma_2\Sigma_1) + m_1^2\Sigma_1^2$,

$$\begin{aligned}\mathcal{Q} = {}& \frac{m_2^2}{6}\,\mathrm{tr}\,\Sigma_2^3 + \frac{m_2^2}{2}\,\mathrm{tr}(\Sigma_1\Sigma_2^2) + \frac{m_1 m_2}{2}\,\mathrm{tr}(\Sigma_1^2\Sigma_2) + \frac{m_1^2}{6}\,\mathrm{tr}\,\Sigma_1^3 \\ & + \frac{1}{2}\,\mathrm{tr}(\Sigma_0\Upsilon_2) + \frac{m_0}{2}\,\mathrm{tr}(\Sigma_0^2\Upsilon_1) + \frac{m_0^2}{6}\,\mathrm{tr}\,\Sigma_0^3.\end{aligned}$$

The sum in (7.23) is the squared-norm integral (4.14) of the path with increments $\Sigma_0, \Sigma_1, \Sigma_2$ and masses m_0, m_1, m_2. In its cross terms $2m_j\, \Sigma_{\le j-1} : \Sigma_j$, the mass is that of the later increment.

Proof. We first determine which products of derivatives of ψ can occur. We then compute the coefficients of the terms that are linear and quadratic in ψ, and finally the term that is cubic in ψ. The computations in Steps 2 and 3 do not use the condition on the diagonal of $\Sigma_{\le 2}$, which enters only at the end of each step.

Step 1: the possible terms. By the remark after (7.18), applied to each of the three steps, $[\mathfrak{E}]_n$ is a sum of products of derivatives of positive order of $\bar\psi$ at 0, of total order $2n$, with coefficients that are polynomials in the entries of the Σ_j and in the m_j. At $\mathbf{y} = 0$, the derivatives of order one vanish, because $\psi'(0) = 0$, the mixed derivatives vanish, because $\bar\psi$ is separable, and $\partial_i^p \bar\psi(0) = \psi^{(p)}(0)$. We conclude that $[\mathfrak{E}]_0 = \bar\psi(0) = 0$ and

$$[\mathfrak{E}]_1 = \mathsf{c}_1\chi, \qquad [\mathfrak{E}]_2 = \mathsf{c}_2\psi^{(4)}(0) + \mathsf{c}_3\chi^2, \qquad [\mathfrak{E}]_3 = \mathsf{c}_4\psi^{(6)}(0) + \mathsf{c}_5\chi\psi^{(4)}(0) + \mathsf{c}_6 w^2 + \mathsf{c}_7\chi^3,$$

where the coefficients c_i are polynomials in the entries of $\Sigma_0, \Sigma_1, \Sigma_2$ (with the masses fixed). These coefficients depend on ψ only through $\psi^{(p)}(0)$ for $2 \le p \le 6$, and we may compute each of them with a convenient choice of ψ.

In Step 2 we take ψ smooth with compact support, for instance the Taylor polynomial of degree six of the given ψ multiplied by a smooth function with compact support that equals one near the origin. We also assume in Step 2 that $\Sigma_0, \Sigma_1, \Sigma_2$ are positive semidefinite. This is no loss of generality: the triples of positive semidefinite matrices contain a nonempty open subset of the triples of symmetric matrices, and an identity between polynomials in the entries that holds on this subset holds for all symmetric triples. For a parameter $\theta \in [-1, 1]$, Lemma 6.9(c), applied to the three steps in turn, shows that the composition of the steps $T_{m_j, \varepsilon\Sigma_j}$ applied to $\theta\bar\psi$ is infinitely differentiable in $(\varepsilon, \theta) \in [0, 1] \times [-1, 1]$. By Lemma 7.6(a) and the chain rule, as in the proof of Lemma 7.6(b), its Taylor coefficients in ε are those of the formal composition. We use independent centered Gaussian vectors $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2$ with covariances $\varepsilon\Sigma_0, \varepsilon\Sigma_1, \varepsilon\Sigma_2$, and we set $\mathbf{S}_j = \mathbf{X}_0 + \cdots + \mathbf{X}_j$, which has covariance $\varepsilon\Sigma_{\le j}$.

Step 2: the terms linear and quadratic in ψ. We replace ψ by $\theta\psi$, so that $\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_4$ are coefficients of θ and $\mathsf{c}_3, \mathsf{c}_5, \mathsf{c}_6$ are coefficients of θ^2. Let $U_3 = \theta\bar\psi$ and $U_j = T_{m_j, \varepsilon\Sigma_j} U_{j+1}$ for $j = 2, 1, 0$, so that $U_0(0)$ is the value of the composition. At $\theta = 0$ all U_j vanish. Differentiating $U_j = m_j^{-1} \log \mathbb{E}_j e^{m_j U_{j+1}}$ in θ at $\theta = 0$, where $\mathbb{E}_j$ is the expectation over $\mathbf{X}_j$, we obtain $\partial_\theta U_j = \mathbb{E}_j \partial_\theta U_{j+1}$ and $\partial_\theta^2 U_j = \mathbb{E}_j \partial_\theta^2 U_{j+1} + m_j \operatorname{Var}_j(\partial_\theta U_{j+1})$. Let $M_j = \mathbb{E}[\bar\psi(\mathbf{S}_2) \mid \mathbf{X}_0, \dots, \mathbf{X}_{j-1}]$, so that $M_0 = \mathbb{E}\bar\psi(\mathbf{S}_2)$ and $M_3 = \bar\psi(\mathbf{S}_2)$. These recursions imply that

$$\partial_\theta U_0(0)\big|_{\theta=0} = \mathbb{E}\bar\psi(\mathbf{S}_2), \qquad \frac{1}{2}\partial_\theta^2 U_0(0)\big|_{\theta=0} = \sum_{j=0}^{2} \frac{m_j}{2}\big(\mathbb{E}M_{j+1}^2 - \mathbb{E}M_j^2\big).$$

The first expression is $\sum_i \mathbb{E}\psi(S_{2,i})$. Its Taylor coefficients in ε are polynomials in the diagonal entries $(\Sigma_{\le 2})_{ii}$ without constant term. These polynomial identities hold for all symmetric triples, and the coefficients vanish under the diagonal condition, which proves that $\mathsf{c}_1 = \mathsf{c}_2 = \mathsf{c}_4 = 0$. For the second, we have $M_{j+1} = (\mathsf{H}\bar{\psi})(\mathbf{S}_j)$, where H is the heat average over the Gaussian vector $\mathbf{X}_{j+1} + \cdots + \mathbf{X}_2$, with covariance $\varepsilon(\Sigma_{\le 2} - \Sigma_{\le j})$. We write $\psi = \psi_2 + \cdots + \psi_6 + \psi_{>6}$, with $\psi_p(y) = \psi^{(p)}(0)y^p/p!$ and $|\psi_{>6}(y)| \le C|y|^7$. All the Gaussian vectors involved have covariances of order ε. The part of M_{j+1} coming from ψ_p with $p \le 6$ is a polynomial in $\mathbf{S}_j$ of the parity of p, and it is $O(\varepsilon^{p/2})$ in every L^q; the part coming from $\psi_{>6}$ is $O(\varepsilon^{7/2})$. A product of the parts coming from ψ_p and $\psi_{p'}$ has expectation $O(\varepsilon^{(p+p')/2})$, with $p = 7$ for $\psi_{>6}$, and this expectation vanishes if $p, p' \le 6$ and $p + p'$ is odd. The terms of order ε^2 and ε^3 of $\mathbb{E}M_{j+1}^2$ come from the three products that we now compute. Let $\Omega = \varepsilon\Sigma_{\le j}$, and let $\Omega' = \varepsilon(\Sigma_{\le 2} - \Sigma_{\le j})$, so that $\Omega_{ii} + \Omega'_{ii} = \varepsilon(\Sigma_{\le 2})_{ii}$. The Gaussian moments in the following three computations are evaluated by the Wick–Isserlis pairing formula [61].

- For ψ_2, we have $M_{j+1} = \frac{\chi}{2}|\mathbf{S}_j|^2 + \text{const}$ and $\operatorname{Var}|\mathbf{S}_j|^2 = 2\|\Omega\|_F^2$. The means $\mathbb{E}M_j$ do not depend on j, and $\mathbb{E}M_{j+1}^2 - \mathbb{E}M_j^2 = \frac{\chi^2\varepsilon^2}{2}\big(\|\Sigma_{\le j}\|_F^2 - \|\Sigma_{\le j-1}\|_F^2\big)$. This gives c_3 and proves (7.23).
- For ψ_3, we have $\mathsf{H}\psi_3(y_i) = \frac{w}{6}(y_i^3 + 3\Omega'_{ii}y_i)$. Using $\mathbb{E}S_iS_l = \Omega_{il}$, $\mathbb{E}S_i^3S_l = 3\Omega_{ii}\Omega_{il}$ and $\mathbb{E}S_i^3S_l^3 = 6\Omega_{il}^3 + 9\Omega_{ii}\Omega_{ll}\Omega_{il}$ for $\mathbf{S} = \mathbf{S}_j$, we find

$$\mathbb{E}M_{j+1}^2 = w^2 \sum_{i,l} \Big\{ \frac{\Omega_{il}^3}{6} + \frac{\Omega_{il}(\Omega_{ii} + \Omega'_{ii})(\Omega_{ll} + \Omega'_{ll})}{4} \Big\}. \tag{7.25}$$

  Under the diagonal condition the second term vanishes, and $\mathbb{E}M_{j+1}^2 = \frac{w^2\varepsilon^3}{6}\operatorname{cub}(\Sigma_{\le j})$. Also $M_0 = 0$, because ψ_3 is odd. This gives c_6.
- For the product of ψ_2 and ψ_4, the means of the two parts of M_j do not depend on j, and by the Wick formula,

$$\operatorname{Cov}\big(M_{j+1}^{(2)}, M_{j+1}^{(4)}\big) = \frac{\chi\psi^{(4)}(0)}{4} \sum_{i,l} \Omega_{il}^2(\Omega_{ll} + \Omega'_{ll}),$$

  where $M_{j+1}^{(p)}$ is the part of M_{j+1} coming from ψ_p. Under the diagonal condition this vanishes, and $\mathsf{c}_5 = 0$.

Step 3: the term cubic in ψ. It remains to compute c_7, and we may take $\psi(y) = \chi y^2/2$. For positive semidefinite covariances and sufficiently small positive ε, the logarithmic determinant below is the Gaussian quadratic-exponential integral, obtained by diagonalizing and applying [107, (A.11)] in each coordinate. We derive its formal extension by computing with the steps directly, for symmetric $\Sigma_0, \Sigma_1, \Sigma_2$. Let Σ and H be symmetric,

$c \in \mathbb{R}$, and $m \in (0,1]$. For a quadratic function $F = c(\varepsilon) + \frac{1}{2}\mathbf{y} \cdot H(\varepsilon)\mathbf{y}$, we have $\Sigma : D^2F = \operatorname{tr}(\Sigma H(\varepsilon))$ and $\nabla F \cdot \Sigma \nabla F = \mathbf{y} \cdot H(\varepsilon)\Sigma H(\varepsilon)\mathbf{y}$. The equation $\partial_\varepsilon F = \frac{1}{2}(\Sigma : D^2F + m\nabla F \cdot \Sigma \nabla F)$ with $F = c + \frac{1}{2}\mathbf{y} \cdot H\mathbf{y}$ at $\varepsilon = 0$ reduces to $c' = \frac{1}{2}\operatorname{tr}(\Sigma H(\varepsilon))$ and $H' = mH(\varepsilon)\Sigma H(\varepsilon)$. Differentiating, we find that the power series $H(\varepsilon) = H(\mathrm{Id} - \varepsilon m\Sigma H)^{-1}$ and $c(\varepsilon) = c - (2m)^{-1}\log\det(\mathrm{Id} - \varepsilon m\Sigma H)$ solve these equations. By the uniqueness of the formal solution,

$$\widehat{T}_{m,\varepsilon\Sigma}\Big(c + \tfrac{1}{2}\mathbf{y} \cdot H\mathbf{y}\Big) = c - \frac{1}{2m}\log\det\big(\mathrm{Id} - \varepsilon m\Sigma H\big) + \tfrac{1}{2}\mathbf{y} \cdot H(\mathrm{Id} - \varepsilon m\Sigma H)^{-1}\mathbf{y},$$

where the right side is expanded in powers of ε. By the substitution rule for formal series, the same holds when c and H are themselves formal power series in ε. Starting from $H_3 = \chi\mathrm{Id}$, the three steps produce the matrices $H_2 = \chi(\mathrm{Id} - \varepsilon m_2\chi\Sigma_2)^{-1}$ and $H_1 = H_2(\mathrm{Id} - \varepsilon m_1\Sigma_1 H_2)^{-1} = \chi\mathrm{Id} + \varepsilon\chi^2\mathrm{Y}_1 + \varepsilon^2\chi^3\mathrm{Y}_2 + O(\varepsilon^3)$, and the value at $\mathbf{y} = 0$ is

$$-\sum_{j=0}^{2}\frac{1}{2m_j}\log\det\big(\mathrm{Id} - \varepsilon m_j\Sigma_j H_{j+1}\big) = \sum_{j=0}^{2}\sum_{p\geq 1}\frac{\varepsilon^p m_j^{p-1}}{2p}\operatorname{tr}\big((\Sigma_j H_{j+1})^p\big).$$

Collecting the terms of order ε^3, we obtain $\chi^3\mathcal{Q}$. For example, for $j = 1$ the terms with $p = 1$, 2, and 3 contribute $\frac{m_2^2}{2}\operatorname{tr}(\Sigma_1\Sigma_2^2)$, $\frac{m_1m_2}{2}\operatorname{tr}(\Sigma_1^2\Sigma_2)$ and $\frac{m_1^2}{6}\operatorname{tr}\Sigma_1^3$, respectively. The terms of order ε and ε^2 are $\frac{\chi}{2}\operatorname{tr}\Sigma_{\leq 2}$, which vanishes under the diagonal condition, and (7.23). This completes the proof. □

In Lemma 7.8 we take $\varepsilon = h$, $\Sigma_j = b\Lambda_j$ and $(m_0, m_1, m_2) = (a/3, a/2, a)$. With these masses $\mathrm{Y}_1 = a(\Sigma_2 + \Sigma_1/2)$ and $\mathrm{Y}_2 = a^2(\Sigma_2 + \Sigma_1/2)^2$. We define the trace polynomial

$$\begin{aligned} C_\chi(k) = {} & \frac{\operatorname{tr}\Lambda_2^3}{6} + \frac{\operatorname{tr}(\Lambda_1\Lambda_2^2)}{2} + \frac{\operatorname{tr}(\Lambda_1^2\Lambda_2)}{4} + \frac{\operatorname{tr}\Lambda_1^3}{24} \\ & + \frac{\operatorname{tr}\big(\Lambda_0(\Lambda_2 + \frac{1}{2}\Lambda_1)^2\big)}{2} + \frac{\operatorname{tr}\big(\Lambda_0^2(\Lambda_2 + \frac{1}{2}\Lambda_1)\big)}{6} + \frac{\operatorname{tr}\Lambda_0^3}{54}, \end{aligned} \tag{7.26}$$

the Wick polynomial

$$\begin{aligned} C_w(k) = \frac{1}{12}\Big\{ & \frac{\mathrm{cub}(\Lambda_0)}{3} + \frac{\mathrm{cub}(\Lambda_0 + \Lambda_1) - \mathrm{cub}(\Lambda_0)}{2} \\ & + \mathrm{cub}(\Lambda_0 + \Lambda_1 + \Lambda_2) - \mathrm{cub}(\Lambda_0 + \Lambda_1)\Big\}, \end{aligned} \tag{7.27}$$

and

$$N_\vartheta^{\mathrm{f}}(k) = \frac{1}{3}\|\Lambda_0\|_F^2 + \frac{1}{2}\big(\|\Lambda_0 + \Lambda_1\|_F^2 - \|\Lambda_0\|_F^2\big) + \|\Lambda_0 + \Lambda_1 + \Lambda_2\|_F^2 - \|\Lambda_0 + \Lambda_1\|_F^2.$$

These are polynomials in k whose coefficients are rational functions of ϑ, and they are even in k by Lemma 7.1(d).

Proposition 7.9 (The constant-mass part). *Let $a = \alpha(q_0)$, $\chi_0 = \chi(q_0, z)$, and $w_0 = w(q_0, z)$. The constant-mass parts of $[V]_0, \dots, [V]_3$ are*

$$0, \qquad 0, \qquad \frac{b^2\chi_0^2}{4}\, a\, N_\vartheta(k), \qquad a^2b^3\chi_0^3\, C_\chi(k) + ab^3 w_0^2\, C_w(k),$$

respectively. Moreover, $C_\chi(0) = C_w(0) = 0$, and $N_\vartheta^{\mathrm{f}} = N_\vartheta$.

Proof. Let $u_0 = u(q_0, z)$ and, for $\mathbf{y} \in \mathbb{R}^3$, write $\varphi(z\mathbf{e} + \mathbf{y}) = 3\Phi(q_0, z) + u_0\, \mathbf{e} \cdot \mathbf{y} + \bar\psi(\mathbf{y})$, where $\bar\psi(\mathbf{y}) = \sum_i \psi(y_i)$ with $\psi(y) = \Phi(q_0, z+y) - \Phi(q_0, z) - u_0 y$. Then $\psi(0) = \psi'(0) = 0$, $\psi''(0) = \chi_0$ and $\psi'''(0) = w_0$. We use Lemma 7.6(e) for each of the three formal steps in (7.22), from the inside out. Formal steps commute with translations, because their coefficients do not depend on $\mathbf{z}$, and the translations produced by the inner steps pass through the outer ones. The total translation is $hbau_0(\frac{1}{3}\Lambda_0 + \frac{1}{2}\Lambda_1 + \Lambda_2)\mathbf{e}$, and the total added constant is $\frac{1}{2}hbau_0^2\, \mathbf{e} \cdot (\frac{1}{3}\Lambda_0 + \frac{1}{2}\Lambda_1 + \Lambda_2)\mathbf{e}$. Both vanish by (7.21), which implies that the value of (7.22) at $z\mathbf{e}$ is $3\Phi(q_0, z)$ plus the value $\mathfrak{E}(h)$ at $\mathbf{y} = 0$ of the three formal steps applied to $\bar\psi$. By (7.21), Lemma 7.8 applies to $\Sigma_j = b\Lambda_j$, and it gives $[\mathfrak{E}]_0 = [\mathfrak{E}]_1 = 0$, $[\mathfrak{E}]_2 = \frac{b^2\chi_0^2}{4} a N_\vartheta^{\mathrm{f}}(k)$ and $[\mathfrak{E}]_3 = a^2b^3\chi_0^3 C_\chi(k) + ab^3 w_0^2 C_w(k)$. The same holds for the baseline, with the values at $k = 0$. By Lemma 7.7, the constant-mass part of $[V]_n$ is the difference between these expressions and their values at $k = 0$.

At $k = 0$ these values vanish. Indeed, at $k = 0$ the composition (7.22) equals φ on the diagonal, for every smooth function Φ_0 in place of $\Phi(q_0, \cdot)$ in the definition of φ. To see this, note that $\Lambda_2 = \mathrm{diag}(-1, -1-\vartheta, -1-\vartheta)$ at $k = 0$, and by Lemma 7.6(d) the inner step maps φ to $\psi_v(z_v) + \psi_x(z_x) + \psi_x(z_y)$, with $\psi_v = \widehat{T}_{a,-bh}\Phi_0$ and $\psi_x = \widehat{T}_{a,-bh(1+\vartheta)}\Phi_0$. The covariance of the shared step is a multiple of $(\mathbf{e}_x + \mathbf{e}_y)(\mathbf{e}_x + \mathbf{e}_y)^{\mathsf{T}}$, and its recursion (7.18) involves only derivatives in the direction $\mathbf{e}_x + \mathbf{e}_y$; on the set $\{z_x = z_y\}$, it is the one-dimensional recursion with variance $bh\vartheta$ and mass $a/2$ applied to $2\psi_x$, which gives $2\widehat{T}_{a,bh\vartheta}\psi_x = 2\widehat{T}_{a,-bh}\Phi_0$ by Lemma 7.6(b) and (c). Similarly, the common step acts on the diagonal as the one-dimensional step with variance bh and mass $a/3$ applied to $\psi_v + 2\widehat{T}_{a,-bh}\Phi_0 = 3\widehat{T}_{a,-bh}\Phi_0$, which gives $3\Phi_0$. The identity holds for every smooth Φ_0 with bounded derivatives of positive order, so that the coefficients χ_0 and w_0 in the expansion above can be prescribed arbitrarily. We conclude that $N_\vartheta^{\mathrm{f}}(0) = C_\chi(0) = C_w(0) = 0$. These values can also be checked directly from the matrices at $k = 0$. For instance, $\|\Lambda_0\|_F^2 = 9$, $\|\Lambda_0 + \Lambda_1\|_F^2 = 5 + 4(1+\vartheta)^2$ and $\|\Lambda_0 + \Lambda_1 + \Lambda_2\|_F^2 = 4 + 2(1+\vartheta)^2$ at $k = 0$, which give $N_\vartheta^{\mathrm{f}}(0) = 3 + 2(1+\vartheta)^2 - 2 - 1 - 2(1+\vartheta)^2 = 0$, and the same computation with cubes in place of squares gives $C_w(0) = 0$. This proves the first claim.

It remains to show that $N_\vartheta^{\mathrm{f}} = N_\vartheta$. Integrating in τ in (7.15) with $i = 0$, we have

$$\sum_{j=1}^{5} \frac{2\kappa_j}{\mathfrak{n}_j} \int_0^1 (\widetilde{Q}_{j-1} + \kappa_j \tau B_j) : B_j \, \mathrm{d}\tau = \sum_{j=1}^{5} \frac{1}{\mathfrak{n}_j} \big(\|\widetilde{Q}_j\|_F^2 - \|\widetilde{Q}_{j-1}\|_F^2 \big).$$

Here $\mathfrak{n}_3 = \mathfrak{n}_4 = \mathfrak{n}_5 = 1$, $\widetilde{Q}_1 = \Lambda_0$, $\widetilde{Q}_2 = \Lambda_0 + \Lambda_1$ and $\widetilde{Q}_5 = \Lambda_0 + \Lambda_1 + \Lambda_2 + \tau_T \mathrm{Id}$. The right side is $N^{\mathrm{f}}_\vartheta(k)$ with Λ_2 replaced by $\Lambda_2 + \tau_T \mathrm{Id}$, and the same holds for the baseline rates. We have $\|U + \tau_T \mathrm{Id}\|_F^2 - \|U\|_F^2 = 2\tau_T \operatorname{tr} U + 3\tau_T^2$ and $\operatorname{tr}(\Lambda_0 + \Lambda_1 + \Lambda_2) = 0$ for both paths. The replacement adds $3\tau_T^2$ to both expressions and does not change their difference. We conclude that $N_\vartheta(k) = N^{\mathrm{f}}_\vartheta(k) - N^{\mathrm{f}}_\vartheta(0) = N^{\mathrm{f}}_\vartheta(k)$. □

The coefficient table

Substituting the matrices (7.20) into (7.26) and (7.27) gives

	k^2	k^4	k^6
$C_\chi(k)$	$\dfrac{2(\vartheta-1)(\vartheta^2+2)}{(\vartheta-2)^2}$	$\dfrac{12\vartheta^3+6\vartheta^2+\vartheta+8}{3(\vartheta-2)^4}$	$\dfrac{(3\vartheta+4)(9\vartheta^2+15\vartheta+10)}{27(\vartheta-2)^6}$
$C_w(k)$	$-\dfrac{(\vartheta-1)(\vartheta^2-\vartheta+2)}{(\vartheta-2)^2}$	$-\dfrac{3\vartheta^3+\vartheta^2+\vartheta+4}{2(\vartheta-2)^4}$	0

(7.28)

These are the coefficients of k^2, k^4 and k^6; the constant terms vanish by Proposition 7.9, and the odd terms vanish because the polynomials are even. The identities can be checked by hand from the explicit matrices, for instance by multiplying by $(\vartheta-2)^6$ and comparing the coefficients of $\vartheta^i k^j$. For $\vartheta = 1/2$, the individual traces and cubic sums are

$$\begin{aligned}
\operatorname{tr}\Lambda_2^3 &= -\tfrac{64}{27}k^6 - \tfrac{32}{3}k^4 - \tfrac{44}{3}k^2 - \tfrac{31}{4},\\
\operatorname{tr}(\Lambda_1\Lambda_2^2) &= \tfrac{64}{81}k^6 + \tfrac{8}{9}k^4 - \tfrac{11}{9}k^2 + \tfrac{9}{4},\\
\operatorname{tr}(\Lambda_1^2\Lambda_2) &= -\tfrac{64}{243}k^6 - \tfrac{40}{81}k^4 - \tfrac{4}{9}k^2 - \tfrac{3}{2},\\
\operatorname{tr}\Lambda_1^3 &= \tfrac{64}{729}k^6 + \tfrac{16}{27}k^4 + \tfrac{4}{3}k^2 + 1,\\
\operatorname{tr}\big(\Lambda_0(\Lambda_2 + \tfrac{1}{2}\Lambda_1)^2\big) &= \tfrac{800}{729}k^6 + \tfrac{488}{81}k^4 + \tfrac{22}{3}k^2 + 3,\\
\operatorname{tr}\big(\Lambda_0^2(\Lambda_2 + \tfrac{1}{2}\Lambda_1)\big) &= -\tfrac{640}{729}k^6 - \tfrac{464}{81}k^4 - 12k^2 - 9,\\
\operatorname{tr}\Lambda_0^3 &= \tfrac{512}{729}k^6 + \tfrac{64}{9}k^4 + 24k^2 + 27,
\end{aligned}$$

and

$$\operatorname{cub}(\Lambda_0) = \tfrac{64}{9}k^4 + 16k^2 + 9, \qquad \operatorname{cub}(\Lambda_0 + \Lambda_1) = \tfrac{88}{9}k^4 + \tfrac{28}{3}k^2 + \tfrac{37}{2},$$
$$\operatorname{cub}(\Lambda_0 + \Lambda_1 + \Lambda_2) = 12k^2 + \tfrac{43}{4}.$$

Inserting these values into (7.26) and (7.27), we obtain

$$C_\chi(k) = -k^2 + \frac{184}{243}k^4 + \frac{6952}{19683}k^6, \qquad C_w(k) = \frac{7}{18}k^2 - \frac{41}{81}k^4 \qquad (\vartheta = 1/2). \tag{7.29}$$

We also checked (7.28), (7.29) and the values of N_ϑ in (7.16) with computer algebra.

7.4 The varying profile

We now compute the part of the Taylor coefficients of V that involves $\alpha'(q_0)$ and $\alpha''(q_0)$. We show that it is exactly the corresponding part of $\frac{b^2\chi(q_0,z)^2}{4}\Delta\Pi$. Together with Proposition 7.9, this gives the pointwise expansion of Proposition 7.12.

The linear part of the terminal function now has a coefficient that changes along the window. We remove it with the following form of (7.19).

Lemma 7.10 (Affine shift along a flow)**.** *Let $(B_j,\lambda_j,m_j)_{j\le N}$ be a stage datum as in Definition 6.1, with continuous masses, let $u\in\mathbb{R}$ and $\mathbf{e}'\in\mathbb{R}^\ell$, and let $f\in\mathcal{C}^\infty$ satisfy $2\ell b\|D^2f\|_\infty\sum_j\lambda_j|B_j|\le 1$. Let F_f and F be the flows with terminal functions f and $u\,\mathbf{e}'\cdot\mathbf{z}+f$. For $\tau'\in(j-1,j)$, write $\lambda(\tau')=\lambda_j$, $B(\tau')=B_j$ and $m(\tau')=m_j(\tau'-j+1)$, and set*

$$\mathbf{s}(\tau)=ub\int_\tau^N\lambda mB\,\mathbf{e}'\,\mathrm{d}\tau',\qquad \bar c(\tau)=\frac{u^2b}{2}\int_\tau^N\lambda m\,\mathbf{e}'\cdot B\mathbf{e}'\,\mathrm{d}\tau'.$$

Then $F(\tau,\mathbf{z})=u\,\mathbf{e}'\cdot\mathbf{z}+\bar c(\tau)+F_f\big(\tau,\mathbf{z}+\mathbf{s}(\tau)\big)$ for $0\le\tau\le N$.

Proof. Let $\widetilde F$ be the right side. By Lemma 6.2(d), F_f is of class $C^{1,2}$ on each closed stage, with bounded first and second spatial derivatives, and solves (6.1); the same properties hold for $\widetilde F$, since $\mathbf{s}$ and $\bar c$ are continuous and continuously differentiable on each closed stage. On a stage, with F_f and its derivatives evaluated at $(\tau,\mathbf{z}+\mathbf{s}(\tau))$,

$$-\partial_\tau\widetilde F=\frac{u^2b\lambda m}{2}\,\mathbf{e}'\cdot B\mathbf{e}'-\partial_\tau F_f+ub\lambda m\,B\mathbf{e}'\cdot\nabla F_f.$$

Since $\nabla\widetilde F=u\mathbf{e}'+\nabla F_f$ and $D^2\widetilde F=D^2F_f$, we have

$$\begin{aligned}\frac{b\lambda}{2}\big\{B:D^2\widetilde F+m\nabla\widetilde F\cdot B\nabla\widetilde F\big\}&=\frac{b\lambda}{2}\big\{B:D^2F_f+m\nabla F_f\cdot B\nabla F_f\big\}\\&\quad+\frac{b\lambda m}{2}\big(u^2\,\mathbf{e}'\cdot B\mathbf{e}'+2u\,\mathbf{e}'\cdot B\nabla F_f\big),\end{aligned}$$

which equals $-\partial_\tau\widetilde F$. We have $\widetilde F(N)=F(N)$, and the terminal function $u\,\mathbf{e}'\cdot\mathbf{z}+f$ has the same Hessian as f. The uniqueness statement of Lemma 6.2(d) implies that $\widetilde F=F$. □

At $\tau=0$ and for constant masses, this is (7.19), applied to each step.

Lemma 7.11 (Contribution of the variation of α)**.** *Fix $q_0\in K$, k with $|k|\le k_0$, and $z\in\mathbb{R}$, and write $\chi_0=\chi(q_0,z)$. For $n\le 3$, the difference between $[V]_n$ and its constant-mass part equals the difference between $\frac{b^2\chi_0^2}{4}[\Delta\Pi]_n$ and its constant-mass part. In other words, it vanishes for $n\le 2$, and for $n=3$ it equals $\frac{b^2\chi_0^2}{4}\alpha'(q_0)N_\vartheta^{(1)}(k)$.*

Proof. The proof has four steps. We remove the linear part of the terminal function by Lemma 7.10, we show that the resulting translations can be replaced by a common one, we list the places where $\alpha'(q_0)$ and $\alpha''(q_0)$ enter, and we compute the only term that survives. Throughout, statements such as "$X = O(h^p)$" mean that $[X]_n = 0$ for $n < p$, and α is regarded as a smooth function on K' whose derivatives at q_0 are free variables, as in the recursion (7.17).

Step 1: removal of the linear part. Let $u_0 = u(q_0, z)$, and write the terminal function as $\sum_i \Phi(T, z_i) = u_0\, \mathbf{e}\cdot\mathbf{z} + \bar\psi_h(\mathbf{z})$. The function $\bar\psi_h$ has bounded derivatives of every positive order and $\|D^2\bar\psi_h\|_\infty \le 1$. By (7.7), Lemma 7.10 applies to the stage datum of (7.9), with $\mathbf{e}' = \mathbf{e}$. At $\tau = 0$, it gives

$$\mathcal{B}(q_0, h, k, \mathbf{z}) = u_0\, \mathbf{e}\cdot\mathbf{z} + \bar c + \widetilde{\mathcal{B}}(\mathbf{z} + \mathbf{s}),$$
$$\mathbf{s} = bu_0 \sum_{j=1}^{5} \frac{1}{\mathfrak{n}_j}\Big(\int_{I_j} \alpha\Big) B_j\mathbf{e}, \qquad \bar c = \frac{bu_0^2}{2} \sum_{j=1}^{5} \frac{1}{\mathfrak{n}_j}\Big(\int_{I_j} \alpha\Big)\mathbf{e}\cdot B_j\mathbf{e},$$

where $\widetilde{\mathcal{B}} = F_{\bar\psi_h}(0,\cdot)$ is the output of the flow with terminal function $\bar\psi_h$, and where we used that $h\kappa_j \int_0^1 m_j = \mathfrak{n}_j^{-1} \int_{I_j} \alpha$. The same holds for the baseline with $\widetilde{\mathcal{B}}^\circ$, $\mathbf{s}^\circ$ and $\bar c^\circ$. By Lemma 7.1(a), $\bar c = \bar c^\circ$. Computing the vectors $B_j\mathbf{e}$ and $B_j^\circ\mathbf{e}$ as in the proof of Lemma 7.1(c), we find

$$\mathbf{s}^\circ = \mathfrak{s}\mathbf{e}, \qquad \mathfrak{s} = bu_0 \int_{q_0}^{T} \alpha, \qquad \mathbf{s} - \mathbf{s}^\circ = bu_0 k\zeta\,(\mathbf{e}_x - \mathbf{e}_y),$$
$$\zeta = \omega_0 J_0 + (\omega_1 - \omega_0) J_1 + (\omega_1 - 2\omega_0) J_2,$$

where $J_0 = \int_{q_0}^{r} \alpha$, $J_1 = \int_r^q \alpha$ and $J_2 = \int_q^{q_0+2h} \alpha$. Subtracting the two representations at $z\mathbf{e}$ and using $\mathcal{B}^\circ(z\mathbf{e}) = 3\Phi(q_0, z)$, we obtain

$$V(q_0, h, k, z) = \widetilde{\mathcal{B}}\big((z+\mathfrak{s})\mathbf{e} + bu_0 k\zeta(\mathbf{e}_x - \mathbf{e}_y)\big) - \widetilde{\mathcal{B}}^\circ\big((z+\mathfrak{s})\mathbf{e}\big). \tag{7.30}$$

Step 2: a common translation. If α were constant on the window, then ζ would vanish, since $\omega_0 + \vartheta(\omega_1 - \omega_0) + (1-\vartheta)(\omega_1 - 2\omega_0) = \omega_1 - (1-\vartheta)\omega_0 = 0$. Since α is Lipschitz on K', we have $|\zeta| \le Ch^2$, so that $\zeta = O(h^2)$. Next, the function $\Delta_{xy}(h, z') = (\mathbf{e}_x - \mathbf{e}_y)\cdot\nabla\widetilde{\mathcal{B}}(z'\mathbf{e})$ satisfies $\Delta_{xy} = O(h^2)$ for every $z' \in \mathbb{R}$. Indeed, at $h = 0$ it equals $\partial_z\Phi(q_0, z') - \partial_z\Phi(q_0, z') = 0$. Its coefficient of order one is the difference between the x and y components of the gradient of

$$\big[\bar\psi_h\big]_1 + \frac{b}{2}\sum_j \kappa_j \Big\{ B_j : D^2\bar\psi_0 + \frac{a}{\mathfrak{n}_j}\nabla\bar\psi_0\cdot B_j\nabla\bar\psi_0 \Big\}$$

at $z'\mathbf{e}$, where $\bar\psi_0 = \bar\psi_h|_{h=0} = \varphi - u_0\,\mathbf{e}\cdot\mathbf{z}$ and $[\bar\psi_h]_1$ is separable and symmetric in the coordinates. At $z'\mathbf{e}$, the gradient of $B_j : D^2\bar\psi_0$ has entries $(B_j)_{ii} w(q_0, z')$, and the gradient

of $\nabla\bar\psi_0 \cdot B_j \nabla\bar\psi_0$ is $2\chi(q_0,z')\gamma\, B_j\mathbf{e}$ with $\gamma = u(q_0,z') - u_0$. These differences vanish after summation, by Lemma 7.1(b) and (c). This proves that $\Delta_{xy} = O(h^2)$. Expanding $\widetilde{\mathcal{B}}$ in the shift $bu_0k\zeta(\mathbf{e}_x - \mathbf{e}_y) = O(h^2)$, we find that the first-order and second-order terms are $O(h^4)$, so that

$$[V]_n = \big[\Delta\widetilde{\mathcal{B}}(z+\mathfrak{s})\big]_n \quad (n \le 3), \qquad \Delta\widetilde{\mathcal{B}}(z') = \widetilde{\mathcal{B}}(z'\mathbf{e}) - \widetilde{\mathcal{B}}^\circ(z'\mathbf{e}).$$

For every z', the function $\Delta\widetilde{\mathcal{B}}(z')$ is $O(h^2)$, with all its derivatives in z': its coefficients of order zero and one vanish by the computation in the proof of Lemma 7.4(ii), with φ replaced by $\bar\psi_0$. Since $\mathfrak{s} = O(h)$, we obtain $[V]_n = [\Delta\widetilde{\mathcal{B}}(z) + \mathfrak{s}\,(\Delta\widetilde{\mathcal{B}})'(z)]_n$ for $n \le 3$.

Step 3: where $\alpha'(q_0)$ and $\alpha''(q_0)$ enter. In $[\Delta\widetilde{\mathcal{B}}(z) + \mathfrak{s}\,(\Delta\widetilde{\mathcal{B}})'(z)]_n$, $n \le 3$, these quantities enter through $\mathfrak{s}$, through the terminal function $\bar\psi_h$, and through the masses.

- The part of $\mathfrak{s}$ that involves them is $bu_0\int_{q_0}^T(\alpha(t)-a)\,\mathrm{d}t = O(h^2)$, and the part of $(\Delta\widetilde{\mathcal{B}})'(z)$ that involves them is $O(h^3)$ by the two items below. Since $(\Delta\widetilde{\mathcal{B}})'(z) = O(h^2)$ and $\mathfrak{s} = O(h)$, neither part contributes to $[\mathfrak{s}\,(\Delta\widetilde{\mathcal{B}})'(z)]_n$ for $n \le 3$.
- In $\bar\psi_h$, the time derivatives of Φ at q_0 involve $\alpha'(q_0)$ only from the second derivative on, and the part of $\bar\psi_h$ that involves $\alpha'(q_0)$ or $\alpha''(q_0)$ is $O(h^2)$. It is separable and symmetric in the coordinates. The first-order change of $\Delta\widetilde{\mathcal{B}}$ under a change $\delta\bar\psi$ of the terminal function of this kind is $O(h^2)\delta\bar\psi$: its coefficients of order zero and one vanish by the computation of Step 2, because $D^2\delta\bar\psi$ is a multiple of Id and $\nabla\delta\bar\psi$ is a multiple of $\mathbf{e}$ on the diagonal. This part contributes $O(h^4)$.
- A mass enters (7.17) only through the gradient term. Each insertion carries a factor h, and the coefficient $[m_j]_l$ of the mass carries the factor h^l and the factor $\alpha^{(l)}(q_0)$. A term that contains $[m_j]_l$ with $l \ge 1$ and only one insertion contains the factor $\nabla\bar\psi_h \cdot B_j\nabla\bar\psi_h$ at $z\mathbf{e}$, because the stages without insertion act as the identity. Since $\nabla\bar\psi_0(z\mathbf{e}) = 0$, we have $\nabla\bar\psi_h(z\mathbf{e}) = O(h)$, so such a term is $O(h^{3+l})$ and does not contribute for $n \le 3$. A term with $[m_j]_2$ and two insertions is $O(h^4)$. The only remaining terms that involve $\alpha'(q_0)$ contain $[m_j]_1$ and exactly two insertions, and they are of order h^3; in them, everything else is evaluated at $h = 0$: the terminal function is $\bar\psi_0$, the point is $z\mathbf{e}$, and the other insertion has the mass $a/\mathfrak{n}_i$.

We note that $\alpha''(q_0)$ does not enter $[V]_n$ for $n \le 3$.

Step 4: the two-insertion term. We use the variable $\tau \in [0,5]$ of (7.9), and we write $\Sigma'(\tau) = bh\kappa_j B_j$ and $m(\tau) = m_j(\tau - j + 1)$ for $\tau \in (j-1, j)$. Two insertions at $\tau_2 < \tau_1$ correspond to an outer insertion at τ_2 and an inner insertion at τ_1. Since $\nabla\bar\psi_0(z\mathbf{e}) = 0$ and $D^2\bar\psi_0(z\mathbf{e}) = \chi_0\mathrm{Id}$, the only term of (7.17) with two insertions that depends on the masses is the heat term of the outer insertion acting on the gradient term of the inner

one:

$$\frac{1}{2}\int_0^5\int_{\tau_2}^5 \Sigma'(\tau_2) : D^2\Big(\frac{m(\tau_1)}{2}\nabla\bar\psi_0\cdot\Sigma'(\tau_1)\nabla\bar\psi_0\Big)(z\mathbf{e})\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 = \frac{\chi_0^2}{2}\int_0^5\int_{\tau_2}^5 m(\tau_1)\,\Sigma'(\tau_1) : \Sigma'(\tau_2)\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2.$$

Here we used that $D^2(\nabla\bar\psi_0\cdot\Sigma'\nabla\bar\psi_0) = 2D^2\bar\psi_0\Sigma' D^2\bar\psi_0$ at $z\mathbf{e}$. The terms in which the outer insertion is a gradient term contain the factor $\nabla\bar\psi_0(z\mathbf{e}) = 0$, and the terms in which both insertions are heat terms do not depend on the masses. Let $\Sigma^{\mathrm{acc}}(\tau) = \int_0^\tau \Sigma'$ be the covariance accumulated from the outer end of the window. Then the right side equals $\frac{\chi_0^2}{2}\int_0^5 m(\tau)\,\Sigma'(\tau) : \Sigma^{\mathrm{acc}}(\tau)\,\mathrm{d}\tau = \frac{\chi_0^2}{4}\int_0^5 m(\tau)\,\partial_\tau\|\Sigma^{\mathrm{acc}}(\tau)\|_F^2\,\mathrm{d}\tau$, which is $\frac{\chi_0^2}{4}$ times the squared-norm integral of the window path ((6.6) with $Q_\star = 0$); this is (7.23) for a continuous profile. By (7.13), its part linear in the coefficients $[m_j]_1$, for the perturbed window minus the baseline window, is $\frac{b^2\chi_0^2}{4}$ times the part of $[\Delta\Pi]_3$ that involves $\alpha'(q_0)$. By (7.15), this part is $\alpha'(q_0)N_\vartheta^{(1)}(k)$. Finally, the part of $[\Delta\Pi]_n$ that involves $\alpha'(q_0)$ or $\alpha''(q_0)$ vanishes for $n \le 2$, which completes the proof. □

Proposition 7.12 (Pointwise expansion of the boundary perturbation)**.** *For $q_0 \in K$, $0 \le h \le h_0$ and $|k| \le k_0$, write $a = \alpha(q_0)$, $\chi_0(z) = \chi(q_0, z)$ and $w_0(z) = w(q_0, z)$. Then*

$$V(q_0,h,k,z) = \frac{b^2\chi_0(z)^2}{4}\,\Delta\Pi(q_0,h,k) + h^3\Big\{a^2b^3\chi_0(z)^3C_\chi(k) + ab^3w_0(z)^2C_w(k)\Big\} + \mathcal{R}(q_0,h,k,z), \tag{7.31}$$

where $|\partial_z^j\mathcal{R}(q_0,h,k,z)| \le Ch^4k^2$ for $j \le 2$ and all z.

The penalty $\Delta\Pi$ in (7.31) is the exact change of the mass-weighted squared-norm integral of the path, with the Parisi distribution function α; it is not replaced by its Taylor polynomial. By the expansion of $\Delta\Pi$ in Section 7.2, it contains the term $\alpha'(q_0)h^3N_\vartheta^{(1)}(k)$. The function V itself depends on $\alpha'(q_0)$ at order h^3. This dependence disappears only after the expectation is taken in Proposition 7.13.

Proof. Let $\mathcal{R}$ be defined by (7.31). Using Proposition 7.9 and Lemma 7.11, together with the expansion $[\Delta\Pi]_2 = aN_\vartheta(k)$, $[\Delta\Pi]_3 = \alpha'(q_0)N_\vartheta^{(1)}(k)$ of Section 7.2, we find that $[\mathcal{R}]_n = 0$ for $n \le 3$, for every k and z. By Lemma 7.4(ii) and (iv), the functions $\partial_h^{\gamma_h}\partial_k^{\gamma_k}\partial_z^j\mathcal{R}$ are bounded for $\gamma_h \le 4$, $\gamma_k \le 2$ and $j \le 2$; for the second and third terms of (7.31) this follows from the bounds on χ and w in Lemma 5.1. Moreover, $\mathcal{R}$ vanishes at $k = 0$ and is even in k, because V, $\Delta\Pi$, C_χ and C_w have these properties. Fix $j \le 2$ and z. The Taylor coefficients of $\partial_k^{\gamma_k}\partial_z^j\mathcal{R}$ of order at most three in h vanish, and Taylor's formula implies that $|\partial_k^{\gamma_k}\partial_z^j\mathcal{R}| \le Ch^4$ for $\gamma_k \le 2$. Since $\partial_z^j\mathcal{R}$ and $\partial_k\partial_z^j\mathcal{R}$ vanish at $k = 0$, we have $\partial_z^j\mathcal{R}(k) = \int_0^k (k-k')\,\partial_k^2\partial_z^j\mathcal{R}(k')\,\mathrm{d}k'$, which implies that $|\partial_z^j\mathcal{R}| \le Ch^4k^2$. □

The identity $[\mathcal{R}]_n = 0$ for $n \le 3$ can also be checked by expanding the five stages of the perturbed path directly, with the derivatives of Φ at (q_0, z), the values $\alpha(q_0)$, $\alpha'(q_0)$, $\alpha''(q_0)$, and k as variables; we did so with computer algebra for several rational values of ϑ.

7.5 The expansion of the trial value

We now take the expectation of (7.31) along the scalar diffusion. Here we use the two stationarity identities. Let

$$M(t) = \alpha(t)^2 \beta^6 \mathbb{E}\chi(t, Y_t)^3, \tag{7.32}$$

which is positive on $(0, q_\beta)$ and infinitely differentiable there by Lemma 5.3. We define the even polynomial

$$P_\vartheta(k) = C_\chi(k) + 2C_w(k) = c_2(\vartheta)k^2 + c_4(\vartheta)k^4 + c_6(\vartheta)k^6.$$

By (7.28), its coefficients are

$$\begin{aligned} c_2(\vartheta) = \frac{2\vartheta(\vartheta-1)}{(\vartheta-2)^2}, \qquad c_4(\vartheta) = \frac{(\vartheta-1)(3\vartheta^2+6\vartheta+4)}{3(\vartheta-2)^4}, \\ c_6(\vartheta) = \frac{(3\vartheta+4)(9\vartheta^2+15\vartheta+10)}{27(\vartheta-2)^6}. \end{aligned} \tag{7.33}$$

For $\vartheta \in (0,1)$ we have $c_2 < 0$, $c_4 < 0$ and $c_6 > 0$. For $\vartheta = 1/2$, by (7.29),

$$P_{1/2}(k) = -\frac{2}{9}k^2 - \frac{62}{243}k^4 + \frac{6952}{19683}k^6. \tag{7.34}$$

The polynomial $-\frac{2}{9}k^2 - \frac{62}{243}k^4 + \frac{6952}{19683}k^6$ is the sum of the trace polynomial (7.26) and twice the Wick polynomial (7.27) at $\vartheta = 1/2$; by Proposition 7.13, it is the coefficient of $M(q_0)h^3$ in the expansion of $\mathcal{T}(q_0, h, k) - 3P_\beta$.

Proposition 7.13 (Local three-replica expansion)**.** *Let $\vartheta \in (0, 1/2]$, $k_0 > 0$, and let $K \subset (0, q_\beta)$ be a compact interval. There exist $h_0 > 0$ and $C < \infty$, depending only on β, ϑ, k_0 and K, such that for all $q_0 \in K$, $0 \le h \le h_0$ and $|k| \le k_0$, the trial value $\mathcal{T}(q_0, h, k)$ of* (7.8) *is an upper bound for $p_{3,n}(D(r, q, \delta))$ whenever this matrix is feasible, and*

$$\left| \mathcal{T}(q_0, h, k) - 3P_\beta - M(q_0)h^3 P_\vartheta(k) \right| \le Ch^4k^2, \tag{7.35}$$

with r, q, δ as in (7.1)*. The same bound holds with $M(q_0)$ replaced by $M(q_0 + h) = M(r)$.*

The constants are uniform in the left endpoint $q_0 \in K$, in $h \le h_0$ and in $|k| \le k_0$; they are not uniform as $\vartheta \downarrow 0$. In terms of $d = \vartheta h$ and $\delta = hk$, the remainder is $Ch^4k^2 = C\delta^2 d^2/\vartheta^2$.

Proof. Let h_0 be as in Lemma 7.4. The first claim is Lemma 7.3(a). Using Corollary 7.5 and Proposition 7.12, we obtain

$$\begin{aligned}\mathcal{T}(q_0,h,k)-3P_\beta &= \frac{b}{4}\big(b\,\mathbb{E}\chi(q_0,Y_{q_0})^2-1\big)\Delta\Pi\\ &\quad + h^3\Big\{a^2b^3\,\mathbb{E}\chi(q_0,Y_{q_0})^3C_\chi(k)+ab^3\,\mathbb{E}w(q_0,Y_{q_0})^2C_w(k)\Big\}\\ &\quad + \mathbb{E}\mathcal{R}+\mathcal{E}_{\mathrm{in}},\end{aligned}$$

with $a=\alpha(q_0)$. Since $q_0\in(0,q_\beta)$, the stationarity identities (5.15) hold at $t=q_0$: we have $b\,\mathbb{E}\chi(q_0,Y_{q_0})^2=1$ and $\mathbb{E}w(q_0,Y_{q_0})^2=2a\,\mathbb{E}\chi(q_0,Y_{q_0})^3$. The first identity cancels the penalty term exactly, and the second turns the braces into $a^2b^3\mathbb{E}\chi(q_0,Y_{q_0})^3\big(C_\chi(k)+2C_w(k)\big)$, so that the second term is $M(q_0)h^3P_\vartheta(k)$. Finally, $|\mathbb{E}\mathcal{R}|\le Ch^4k^2$ by Proposition 7.12, and $|\mathcal{E}_{\mathrm{in}}|\le Ch^4k^4\le Ck_0^2h^4k^2$ by Corollary 7.5. This proves (7.35).

For the last claim, recall that M is continuously differentiable on K'. We have $|M(q_0+h)-M(q_0)|\le Ch$ and $|P_\vartheta(k)|\le Ck^2$ for $|k|\le k_0$, and replacing $M(q_0)$ by $M(q_0+h)$ changes the left side of (7.35) by at most Ch^4k^2. □

7.6 A uniform cubic deficit on a band

We now express the expansion in the variables $d=q-r$ and δ, and we bound the trial value on a band where both are comparable to a small scale ρ. With $h=d/\vartheta$ and $k=\vartheta\delta/d$, we have $-h^3P_\vartheta(k)=\mathcal{G}_\vartheta(d,|\delta|)$, where

$$\mathcal{G}_\vartheta(d,x)=-\frac{c_2}{\vartheta}\,x^2d-\vartheta c_4\frac{x^4}{d}-\vartheta^3c_6\frac{x^6}{d^3}\qquad(d,x>0),\tag{7.36}$$

and $c_j=c_j(\vartheta)$. Note that $-c_2/\vartheta=2(1-\vartheta)/(2-\vartheta)^2>0$. For $E>0$ we set

$$G(\vartheta,E)=-\frac{c_2}{\vartheta}-\vartheta c_4E^2-\vartheta^3c_6E^4,\qquad \mu_1=-\frac{c_2}{\vartheta}+\vartheta c_4(2E)^2,\qquad \mu_2=-\frac{c_2}{\vartheta}-3\vartheta^3c_6(2E)^4,$$

so that $\mathcal{G}_\vartheta(\rho,E\rho)=E^2G(\vartheta,E)\rho^3$. Since $c_4<0<c_6$, we have $G(\vartheta,E)\ge\mu_2$.

Lemma 7.14 (Monotonicity on the band)**.** *Let $\vartheta\in(0,1/2]$ and $E>0$ satisfy*

$$-\frac{c_2(\vartheta)}{\vartheta}+\vartheta c_4(\vartheta)(2E)^2>0,\qquad -\frac{c_2(\vartheta)}{\vartheta}-3\vartheta^3c_6(\vartheta)(2E)^4>0,\tag{7.37}$$

that is, $\mu_1>0$ and $\mu_2>0$. Let $0\le\sigma<\min\{\mu_1,\mu_2\}$ and $\rho>0$. Then the function $(d,x)\mapsto\mathcal{G}_\vartheta(d,x)-\sigma x^2d$ is increasing in each variable on the band $\{d\ge\rho,\ E\rho\le x\le 2E\rho\}$, and its minimum over the band is $E^2(G(\vartheta,E)-\sigma)\rho^3$, attained at $(\rho,E\rho)$.

Proof. Differentiating (7.36), we obtain

$$\frac{\partial_d\mathcal{G}_\vartheta}{x^2}=-\frac{c_2}{\vartheta}+\vartheta c_4\Big(\frac{x}{d}\Big)^2+3\vartheta^3c_6\Big(\frac{x}{d}\Big)^4,\qquad \frac{\partial_x\mathcal{G}_\vartheta}{2xd}=-\frac{c_2}{\vartheta}-2\vartheta c_4\Big(\frac{x}{d}\Big)^2-3\vartheta^3c_6\Big(\frac{x}{d}\Big)^4.$$

On the band, $0 < x/d \le 2E\rho/\rho = 2E$. Since $c_4 < 0 < c_6$, the first expression is at least μ_1 and the second is at least μ_2. Subtracting $\sigma x^2 d$ subtracts σ from both normalized derivatives, and the partial derivatives of $\mathcal{G}_\vartheta - \sigma x^2 d$ in d and x are at least $(\mu_1 - \sigma)x^2 > 0$ and $2(\mu_2 - \sigma)xd > 0$. The band is the product $[\rho, \infty) \times [E\rho, 2E\rho]$, and so the minimum is attained at the corner $(\rho, E\rho)$, where the value is $\mathcal{G}_\vartheta(\rho, E\rho) - \sigma E^2\rho^3 = E^2(G(\vartheta, E) - \sigma)\rho^3$. □

The lemma bounds $\mathcal{G}_\vartheta - \sigma x^2 d$ below by its value at one corner of the band. It does not say that $G(\vartheta, E)$ multiplies $x^2 d$ throughout the band.

Proposition 7.15 (A uniform cubic deficit)**.** *Let $I \subset (0, q_\beta)$ be a nonempty compact interval, and let $m_\beta > 0$ satisfy $m_\beta \le \inf_{r\in I} M(r)$, for instance the constant of Corollary 5.5 for an interval $I_\beta \supset I$. Let $\vartheta \in (0, 1/2]$ and $E > 0$ satisfy* (7.37)*, and let $0 < \sigma < \min\{\mu_1, \mu_2\}$. There exists $d_0 > 0$, depending only on β, I, ϑ, E and σ, such that the following holds. Let $0 < \rho \le d_0$, and let $r \in I$, d and δ satisfy*

$$\rho \le d \le d_0, \qquad E\rho \le |\delta| \le 2E\rho.$$

Set $h = d/\vartheta$, $q_0 = r - h$ and $k = \vartheta\delta/d$. Then $\mathcal{T} = \mathcal{T}(q_0, h, k)$ is an upper bound for $p_{3,n}(D(r, r+d, \delta))$ whenever this matrix is feasible, and

$$\mathcal{T} \le 3P_\beta - m_\beta\{\mathcal{G}_\vartheta(d, |\delta|) - \sigma\delta^2 d\} \le 3P_\beta - m_\beta E^2\big(G(\vartheta, E) - \sigma\big)\rho^3.$$

In particular, if $0 < \bar\sigma < G(\vartheta, E)$ and $\sigma = \min\{\bar\sigma, \mu_1, \mu_2\}/2$, then

$$\mathcal{T} \le 3P_\beta - m_\beta E^2\big\{G(\vartheta, E) - \bar\sigma\big\}\rho^3. \tag{7.38}$$

The parameters are chosen in the order of Table 7.3, all before ρ, and before n in the later sections. The deficit (7.38) is uniform over the band, but it is a bound by the value at the corner of the band; at other points of the band the deficit $m_\beta\{\mathcal{G}_\vartheta(d, |\delta|) - \sigma\delta^2 d\}$ may be much larger. A fixed $d = d_0$ does not make the remainder of (7.35) small compared with $\delta^2 d$; it is the choice of d_0 in the proof that makes it so.

Proof. We first fix the constants, and then combine Proposition 7.13 with Lemma 7.14.

Step 1: constants. Let $k_0 = 2\vartheta E$ and $K = [\frac{1}{2}\min I, \max I]$, a compact subinterval of $(0, q_\beta)$. Let h_0 and C be the constants of Proposition 7.13 for ϑ, k_0 and K. By Lemma 5.3, the function M is continuous and positive on the compact interval I, and we set

$$\underline{M}_I = \min_{r\in I} M(r) > 0, \qquad d_0 = \min\Big\{\vartheta h_0,\ \frac{\vartheta}{2}\min I,\ \frac{\underline{M}_I \sigma\vartheta^2}{C}\Big\}.$$

These constants depend only on β, I, ϑ, E and σ.

Parameter	Chosen in terms of	Purpose
I, m_β	fixed first	centers; scalar factor $M \geq m_\beta$ on I
ϑ, E	(7.37)	geometry of the band
$k_0 = 2\vartheta E$, K	ϑ, E, I	range of k and of q_0
h_0, C	β, ϑ, k_0, K	Proposition 7.13
σ (or $\bar\sigma$)	μ_1, μ_2 (and $G(\vartheta, E)$)	slack in Lemma 7.14
d_0	h_0, C, $\underline{M}_I$, σ, I	$h \leq h_0$, $q_0 \in K$, absorption of the remainder
ρ	$0 < \rho \leq d_0$	scale of the band

Table 7.3. Order of the choices in Proposition 7.15.

Step 2: the expansion. Let ρ, r, d, δ be as in the statement. Then $h = d/\vartheta \leq h_0$, $q_0 = r - h \geq \min I - \frac{1}{2}\min I$, so that $q_0 \in K$, and $|k| = \vartheta|\delta|/d \leq 2\vartheta E\rho/\rho = k_0$. With these parameters, $q_0 + h = r$, $q_0 + (1+\vartheta)h = r + d$ and $hk = \delta$. The endpoint of the path is $D(r, r+d, \delta)$, and the first claim is Proposition 7.13. By (7.35) with $M(r)$,

$$\mathcal{T} \leq 3P_\beta - M(r)\,\mathcal{G}_\vartheta(d, |\delta|) + \frac{C}{\vartheta^2}\delta^2 d^2.$$

Since $d \leq d_0$ and $r \in I$, we have

$$\frac{C}{\vartheta^2}\delta^2 d^2 \leq \underline{M}_I \sigma \delta^2 d \leq M(r)\sigma\delta^2 d.$$

By Lemma 7.14, the braces in the next display are at least $E^2(G(\vartheta, E) - \sigma)\rho^3$, which is positive because $G(\vartheta, E) \geq \mu_2 > \sigma$. Since $M(r) \geq m_\beta$, we obtain

$$\mathcal{T} \leq 3P_\beta - M(r)\big\{\mathcal{G}_\vartheta(d, |\delta|) - \sigma\delta^2 d\big\} \leq 3P_\beta - m_\beta\big\{\mathcal{G}_\vartheta(d, |\delta|) - \sigma\delta^2 d\big\},$$

and the lower bound on the braces gives the second inequality of the proposition. For the last claim, $\sigma \leq \bar\sigma/2 < \bar\sigma$. □

An explicit admissible choice. Let $\vartheta = 1/2$ and $E = 1/100$. By (7.34), $-c_2/\vartheta = 4/9$, $c_4 = -62/243$ and $c_6 = 6952/19683$. These values satisfy $-1 < c_4 < 0 < c_6 < 1$, which implies that

$$\mu_1 = \frac{4}{9} + 2c_4E^2 > \frac{4}{9} - 2E^2 > 0, \qquad \mu_2 = \frac{4}{9} - 6c_6E^4 > \frac{4}{9} - 6E^4 > 0,$$

so (7.37) holds, and

$$G\Big(\frac{1}{2}, E\Big) = \frac{4}{9} + \frac{31}{243}E^2 - \frac{869}{19683}E^4 > \frac{4}{9} - \frac{E^4}{8} > \frac{1}{3}.$$

The exact values are $\mu_1 = 269969/607500$, $\mu_2 = 18224999131/41006250000$ and $G(1/2, 1/100) = 874825109131/1968300000000$. This choice is used in Section 10. It is compatible with every fixed $\beta > 1$: the constants of Proposition 7.13 depend on β, but (7.37) does not.

Scope of the local construction. Proposition 7.15 concerns only matrices with $q - r \le d_0$. For $q - r > d_0$, Section 9 uses the quenched comparison with a fixed width of Lemma 9.5, which does not require the regularity of α near q. The bound of Proposition 7.15 is a bound on the quenched pressure. To bound the positive fractional moments for the probability estimates, we use the separate interpolation of Section 8, which uses the trial path of this section with the profile changed on $[0, q_0)$; by the remark after Lemma 7.3, this change does not affect V or $\Delta\Pi$.

SECTION 8

Fractional moments of constrained partition functions

In this section we bound positive fractional moments of the partition function of three configurations with prescribed overlaps. The main result is Lemma 8.7. It states that, for every overlap matrix D in the local band of Section 7, the normalized fractional moment $(ns)^{-1}\log \mathbb{E}Z_{3,n}(D)^s$ of order $s = S\rho$ is at most $3P_\beta - g\rho^3$, uniformly in n. In Section 9 we use this bound and Markov's inequality to obtain a single disorder event on which all these constrained partition functions are small. That event is the static input of the dynamical argument in Sections 10 to 12.

We first explain why a fractional moment is needed. The cubic deficit of Section 7 bounds the quenched constrained pressure $p_{3,n}(D) = n^{-1}\mathbb{E}\log Z_{3,n}(D)$ by $3P_\beta$ minus a multiple of ρ^3. Gaussian concentration turns a mean deficit of order ρ^3 into a probability bound $\exp(-cn\rho^6)$, and for $\rho \asymp n^{-7/30}$ the exponent $n\rho^6 \asymp n^{-2/5}$ tends to zero. With a bound of the same order on the moment of order $s \asymp \rho$, we obtain instead the probability bound $\exp(-cns\rho^3) = \exp(-cn\rho^4)$, and $n\rho^4 \asymp n^{1/15}$. The proof has three ingredients.

1. Proposition 8.3 is a matrix interpolation bound for the moment of order s of a constrained partition function. Its right side is the trial value of the quenched comparison Proposition 4.19, which is [77, Proposition 2.3]. However, the comparison path must have all its masses at least s, and the left side carries the correction $\beta^2\ell^2 s/(4n)$ produced by the Gaussian completion of Lemma 3.1.
2. The three-replica path of Section 7 has masses $\alpha/3$ near the origin, and $\alpha(t) \to 0$ as $t \to 0$. We replace α by the profile $\max\{\alpha, 3s\}$. Its cost is $O(s^5)$ by the estimate of Aronow and Lopatto recalled in Proposition 4.12, and Lemma 8.6 controls the effect of this change on the local covariance perturbation, which starts at a time bounded away from zero.
3. Lemma 8.7 combines these estimates with the uniform cubic deficit (7.38).

The fractional comparison and the local fractional bound are new relative to [77]. The comparison adapts the quenched bound [77, Proposition 2.3], and the positive-moment mechanism and the profile $\max\{\alpha, 3s\}$ come from [8]; see Remark 8.4.

Section 8.1 defines Gaussian hierarchies whose outermost integration has positive mass and computes their variance derivatives. Section 8.2 proves Proposition 8.3. Sections 8.3 and 8.4 treat the profile $\max\{\alpha, 3s\}$, and Section 8.5 proves Lemma 8.7.

Standing notation. Throughout this section and the next, $\beta > 1$ is fixed, and $\alpha = \alpha_\beta$, μ_β, q_β, Φ, u, χ, Y and Γ are as in Section 4. Recall that $\Gamma(t) = t$ for $0 \le t \le q_\beta$ and that α is smooth on $[0, q_\beta)$, by Corollary 4.8. We use the constrained partition functions $Z_{\ell,n}(D)$ and pressures $p_{\ell,n}(D)$ of Definition 4.16, the matrices $D_2(q)$ and $D(r,q,\delta)$ of (4.16), the integration steps $T_{m,\Sigma}$ of Definition 4.1, and the admissible paths and trial values $\mathcal{T}_D(Q,m,\lambda)$ of Definitions 4.14 and 4.15. For real $\ell \times \ell$ matrices U, U', not necessarily symmetric, $U : U' = \sum_{a,a'} U_{aa'} U'_{aa'}$ and $\|U\|_F^2 = U : U$. For a feasible $\ell \times \ell$ matrix D and $0 < s \le 1$, we define the normalized fractional moment

$$p_{\ell,n}^{(s)}(D) = \frac{1}{ns} \log \mathbb{E} Z_{\ell,n}(D)^s. \tag{8.1}$$

By Jensen's inequality, we have $p_{\ell,n}(D) \le p_{\ell,n}^{(s)}(D)$, so an upper bound on $p_{\ell,n}^{(s)}(D)$ is stronger than the same upper bound on $p_{\ell,n}(D)$. Throughout, $s \in (0,1]$ denotes the order of a fractional moment; in Section 8.3 it is also the parameter of the profile $\max\{\alpha, 3s\}$ of Proposition 4.12.

8.1 Hierarchies with a positive outer mass

In the hierarchies of Proposition 4.18, the quenched randomness is integrated by an ordinary expectation, which corresponds to mass zero. To control a fractional moment of order s, we instead integrate it at mass s.

Definition 8.1 (Hierarchy with positive outer mass)**.** Let $\mathcal{S}$ be a nonempty finite set, let $N \ge 1$, let $s \in (0,1]$, and let $s \le m_1 \le \cdots \le m_N \le 1$. Let $\omega, \mathbf{g}_1, \dots, \mathbf{g}_N$ be independent standard Gaussian vectors in finite-dimensional spaces, and for $x \in \mathcal{S}$ let $U(x) = U(x; \omega, \mathbf{g}_1, \dots, \mathbf{g}_N)$ be affine in $(\omega, \mathbf{g}_1, \dots, \mathbf{g}_N)$. We set

$$\Psi_N = \log \sum_{x \in \mathcal{S}} e^{U(x)}, \qquad \Psi_{j-1} = \frac{1}{m_j} \log \mathbb{E}_j e^{m_j \Psi_j} \quad (j = N, \dots, 1), \qquad \Psi_{-1} = \frac{1}{s} \log \mathbb{E}_\omega e^{s \Psi_0},$$

where $\mathbb{E}_j$ integrates $\mathbf{g}_j$ and $\mathbb{E}_\omega$ integrates ω. Then Ψ_j is a function of $(\omega, \mathbf{g}_1, \dots, \mathbf{g}_j)$, and the number Ψ_{-1} is the value of the hierarchy. The path law is the law of $(\omega, \mathbf{g}_1, \dots, \mathbf{g}_N, \mathsf{x})$ under which ω has density $e^{s(\Psi_0 - \Psi_{-1})}$ with respect to its Gaussian law, given $(\omega, \mathbf{g}_1, \dots, \mathbf{g}_{j-1})$ the vector $\mathbf{g}_j$ has density $e^{m_j(\Psi_j - \Psi_{j-1})}$, and given all Gaussian variables the configuration $\mathsf{x} \in \mathcal{S}$ has the Gibbs law proportional to $e^{U(\mathsf{x})}$. For $0 \le l \le N$, $\mathbb{E}_l^{(2)}$ is the expectation under the pair law of $(\mathsf{x}^1, \mathsf{x}^2)$ sharing $(\omega, \mathbf{g}_1, \dots, \mathbf{g}_l)$, with conditionally independent continuations. We set $m_0 = s$, $m_{N+1} = 1$ and $\Delta m_l = m_{l+1} - m_l$ for $0 \le l \le N$.

Note that the recursion normalizes the densities, and that all integrals are finite since Ψ_N is Lipschitz in the Gaussian variables. As $s \to 0$, the outermost step becomes

$\Psi_{-1} = \mathbb{E}_\omega \Psi_0$, the quenched root of Proposition 4.18, where $m_0 = 0$. The next lemma is the analogue of Proposition 4.18, which is [77, Lemma 2.2]; the only change is that the weights now sum to $1 - s$.

Lemma 8.2 (Variance derivative)**.** *In the setting of Definition 8.1, suppose that a standard Gaussian vector* $\mathbf{g}$*, which is either* $\mathbf{g}_j$ *for some* $1 \le j \le N$ *or a component of* ω *(in which case* $j = 0$*), enters* U *only through an additive term* $\sqrt{v}\,\mathbf{g} \cdot e(x)$*, where* e *is a fixed vector-valued function on* $\mathcal{S}$*. Then, with all other parameters fixed, for* $v > 0$*,*

$$\frac{\partial \Psi_{-1}}{\partial v} = \frac{1}{2}\Big(\mathbb{E}\|e(\mathsf{x})\|^2 - \sum_{l=j}^{N} \Delta m_l\, \mathbb{E}_l^{(2)}\, e(\mathsf{x}^1) \cdot e(\mathsf{x}^2)\Big), \tag{8.2}$$

where $\mathbb{E}$ *is the expectation under the path law. The weights* $\Delta m_0, \dots, \Delta m_N$ *are nonnegative and sum to* $1 - s$*.*

Proof. We follow the proof of [77, Lemma 2.2] and write it out, because the outermost level is now integrated at positive mass. The proof has three steps: a recursion for Hessians in an auxiliary deterministic field, the variance derivative at level j, and a telescoping sum.

Step 1: the Hessian recursion. For a deterministic vector $\mathbf{y}$, let $\Psi_l(\mathbf{y})$ be the values of the hierarchy when $\mathbf{y} \cdot e(x)$ is added to $U(x)$. For $0 \le l \le N$, we write $\bar{e}_l = \mathbb{E}[e(\mathsf{x}) \mid \omega, \mathbf{g}_1, \dots, \mathbf{g}_l]$ under the path law, and let H_l be the Hessian of Ψ_l in $\mathbf{y}$ at $\mathbf{y} = 0$. Note that $\nabla_{\mathbf{y}} \Psi_N = \bar{e}_N$ and that $H_N = \mathrm{Cov}(e(\mathsf{x}) \mid \omega, \mathbf{g}_1, \dots, \mathbf{g}_N)$, the covariance under the Gibbs law. For $1 \le l \le N$, differentiating $\Psi_{l-1} = m_l^{-1} \log \mathbb{E}_l e^{m_l \Psi_l}$ once and twice, we obtain

$$\nabla_{\mathbf{y}} \Psi_{l-1} = \bar{e}_{l-1}, \qquad H_{l-1} = \mathbb{E}[H_l \mid \omega, \mathbf{g}_1, \dots, \mathbf{g}_{l-1}] + m_l\, \mathrm{Cov}(\bar{e}_l \mid \omega, \mathbf{g}_1, \dots, \mathbf{g}_{l-1}),$$

where for the first identity we used the tower property of conditional expectation. Taking traces and expectations, we find

$$\mathbb{E}\,\mathrm{tr}\, H_N = \mathbb{E}\|e(\mathsf{x})\|^2 - \mathbb{E}\|\bar{e}_N\|^2, \qquad \mathbb{E}\,\mathrm{tr}\, H_{l-1} = \mathbb{E}\,\mathrm{tr}\, H_l + m_l\big(\mathbb{E}\|\bar{e}_l\|^2 - \mathbb{E}\|\bar{e}_{l-1}\|^2\big). \tag{8.3}$$

Step 2: the derivative at level j. The vector $\mathbf{g}$ enters Ψ_j only through the shift $\mathbf{y} = \sqrt{v}\,\mathbf{g}$. We use the heat equation $\partial_v \mathbb{E} f(\sqrt{v}\,\mathbf{g}) = \frac{1}{2}\mathbb{E}\Delta f(\sqrt{v}\,\mathbf{g})$, valid for smooth f of moderate growth. Applying it to $f = e^{m_j \Psi_j}$ in the step that integrates level j, at mass m_j (with $m_0 = s$ for ω), we obtain

$$\partial_v \Psi_{j-1} = \frac{1}{2}\mathbb{E}\big[\mathrm{tr}\, H_j + m_j \|\bar{e}_j\|^2 \;\big|\; \omega, \mathbf{g}_1, \dots, \mathbf{g}_{j-1}\big],$$

which for $j = 0$ is to be read without conditioning. At the earlier levels, differentiating $m^{-1} \log \mathbb{E} e^{m\Psi}$ in v gives the tilted expectation of $\partial_v \Psi$; this holds at every level, including the outermost one at mass s. Iterating, we conclude that

$$2\,\partial_v \Psi_{-1} = \mathbb{E}\,\mathrm{tr}\, H_j + m_j \mathbb{E}\|\bar{e}_j\|^2. \tag{8.4}$$

Step 3: telescoping. Inserting (8.3) for $l = N, \dots, j+1$ into (8.4), we have

$$2\,\partial_v \Psi_{-1} = \mathbb{E}\|e(\mathsf{x})\|^2 - \mathbb{E}\|\bar{e}_N\|^2 + \sum_{l=j+1}^{N} m_l \big(\mathbb{E}\|\bar{e}_l\|^2 - \mathbb{E}\|\bar{e}_{l-1}\|^2\big) + m_j \mathbb{E}\|\bar{e}_j\|^2.$$

For $j \le l \le N$, the coefficient of $\mathbb{E}\|\bar{e}_l\|^2$ is $m_l - m_{l+1} = -\Delta m_l$, where we used $m_{N+1} = 1$. Two paths that share $(\omega, \mathbf{g}_1, \dots, \mathbf{g}_l)$ are conditionally independent, which implies that $\mathbb{E}\|\bar{e}_l\|^2 = \mathbb{E}_l^{(2)} e(\mathsf{x}^1) \cdot e(\mathsf{x}^2)$. This proves (8.2). The weights are nonnegative because the masses are ordered, and they sum to $m_{N+1} - m_0 = 1 - s$.

All differentiations under the integral sign are justified by dominated convergence. Indeed, $\mathcal{S}$ and the number of levels are finite, the gradients of U in the Gaussian variables are bounded, and the values grow at most linearly in these variables. □

In the quenched case $m_0 = 0$, the term $m_0 \mathbb{E}\|\bar{e}_0\|^2$ is absent and the weights sum to one; this is the only difference from [77, Lemma 2.2]. The same term appears in the differentiation of a positive fractional moment of the one-replica partition function by Aronow and Lopatto [8, Section 2.3, (2.10)–(2.11)], where it removes the prescribed mass at zero from the integral against the variational measure.

8.2 The fractional matrix comparison

Proposition 8.3 (Fractional matrix comparison)**.** *Let $n \ge 2$, $\ell \in \{2, 3\}$ and $s \in (0, 1]$, let D be a feasible $\ell \times \ell$ matrix, let (Q, m) be an admissible path of $\ell \times \ell$ matrices whose masses satisfy $m_1 \ge s$, and let λ be a symmetric $\ell \times \ell$ matrix. Then*

$$p_{\ell,n}^{(s)}(D) + \frac{\beta^2 \ell^2 s}{4n} \le \mathcal{T}_D(Q, m, \lambda) = \Psi(Q, m, \lambda) - \lambda : D + \frac{\beta^2}{4}\Big(\|D - Q_N\|_F^2 - \int m \,\mathrm{d}\|Q\|_F^2\Big). \tag{8.5}$$

The same inequality holds for every trial value for D, in the sense of Definition 4.15, that is a limit along admissible paths whose masses are all at least s.

The right side of (8.5) is exactly the right side of the quenched comparison Proposition 4.19. Every trial bound for $p_{\ell,n}(D)$ obtained from a path whose masses are all at least s also bounds $p_{\ell,n}^{(s)}(D) + \beta^2 \ell^2 s/(4n)$.

Proof. We interpolate between the completed Hamiltonian and independent fields at the n sites, as in the proof of [77, Proposition 2.3]. The difference is that the Hamiltonian disorder is integrated at the outer mass s instead of by an ordinary expectation. The proof has four steps: the covariances, the derivative of the interpolation, its endpoints, and the effect of the completion.

Step 1: covariances. Let J_0 be a standard Gaussian variable independent of J, set $G = J_0/\sqrt{2}$, and let $H_n^c = H_n + G$ be the completed Hamiltonian of Lemma 3.1, so that $\mathbb{E}H_n^c(x)H_n^c(y) = nR(x,y)^2/2$. For two tuples $\boldsymbol{x} = (x^1, \dots, x^\ell)$ and $\boldsymbol{y}$ in Σ_n^ℓ, let $R(\boldsymbol{x}, \boldsymbol{y})$ be the cross-overlap matrix with entries $R(x^a, y^{a'})$; it need not be symmetric. Then

$$\mathbb{E}\Big[\beta \sum_{a=1}^{\ell} H_n^c(x^a) \cdot \beta \sum_{a'=1}^{\ell} H_n^c(y^{a'})\Big] = \frac{n\beta^2}{2} \sum_{a,a'=1}^{\ell} R(x^a, y^{a'})^2 = \frac{n\beta^2}{2} \|R(\boldsymbol{x}, \boldsymbol{y})\|_F^2. \tag{8.6}$$

The process $\beta \sum_a H_n^c(x^a)$ is a linear form in the standard Gaussian vector $\omega = ((J_{ij})_{i<j}, J_0)$, as required in Lemma 8.2. For the site fields, let $\mathbf{g}_{j,i}$, $1 \le j \le N$, $1 \le i \le n$, be independent standard Gaussian vectors in $\mathbb{R}^\ell$, let $\mathbf{g}_j = (\mathbf{g}_{j,i})_{i \le n}$, and write $\boldsymbol{x}_i = (x_i^1, \dots, x_i^\ell) \in \{-1, 1\}^\ell$ for the spins of $\boldsymbol{x}$ at site i. With $\Delta Q_j = Q_j - Q_{j-1}$, we have

$$\mathbb{E}\Big[\sum_{i=1}^{n} \beta \Delta Q_j^{1/2} \mathbf{g}_{j,i} \cdot \boldsymbol{x}_i \sum_{i=1}^{n} \beta \Delta Q_j^{1/2} \mathbf{g}_{j,i} \cdot \boldsymbol{y}_i\Big] = \beta^2 \sum_{i=1}^{n} \boldsymbol{x}_i^{\mathsf{T}} \Delta Q_j \boldsymbol{y}_i = n\beta^2 \, \Delta Q_j : R(\boldsymbol{x}, \boldsymbol{y}). \tag{8.7}$$

Step 2: the interpolation. For $0 \le t \le 1$, consider the hierarchy of Definition 8.1 with $\mathcal{S} = \Sigma_D = \{\boldsymbol{x} \in \Sigma_n^\ell : R(\boldsymbol{x}, \boldsymbol{x}) = D\}$, the masses $s \le m_1 \le \cdots \le m_N$, the Gaussian vectors $\omega, \mathbf{g}_1, \dots, \mathbf{g}_N$, and

$$U_t(\boldsymbol{x}) = \sqrt{t}\,\beta \sum_{a=1}^{\ell} H_n^c(x^a) + \sqrt{1-t} \sum_{i=1}^{n} \sum_{j=1}^{N} \beta \Delta Q_j^{1/2} \mathbf{g}_{j,i} \cdot \boldsymbol{x}_i + \sum_{i=1}^{n} \boldsymbol{x}_i^{\mathsf{T}} \lambda \boldsymbol{x}_i.$$

Let $\Theta(t)$ be its value divided by n. The Hamiltonian enters at level 0 with variance parameter t, and the site fields of level j enter with variance parameter $1 - t$. For $\boldsymbol{x} \in \Sigma_D$ we have $R(\boldsymbol{x}, \boldsymbol{x}) = D$. Applying Lemma 8.2 to each of these Gaussian vectors and using (8.6) and (8.7), we obtain, for $0 < t < 1$,

$$\begin{aligned}
\Theta'(t) &= \frac{1}{2n}\Big[\frac{n\beta^2}{2}\Big(\|D\|_F^2 - \sum_{l=0}^{N} \Delta m_l \mathbb{E}_l^{(2)} \|R\|_F^2\Big) \\
&\qquad - \sum_{j=1}^{N} n\beta^2 \Big(\Delta Q_j : D - \sum_{l=j}^{N} \Delta m_l \mathbb{E}_l^{(2)} \, \Delta Q_j : R\Big)\Big] \\
&= \frac{\beta^2}{4}\Big[\|D\|_F^2 - 2Q_N : D - \sum_{l=0}^{N} \Delta m_l \mathbb{E}_l^{(2)} \big(\|R\|_F^2 - 2Q_l : R\big)\Big],
\end{aligned}$$

where $R = R(\boldsymbol{x}^1, \boldsymbol{x}^2)$ is the cross-overlap matrix of the terminal tuples of the two paths, and in the last line we summed ΔQ_j over $j \le l$, using $Q_0 = 0$. For every matrix R we have $\|R\|_F^2 - 2Q_l : R = \|R - Q_l\|_F^2 - \|Q_l\|_F^2$. Since $\Delta m_l \ge 0$, this implies that

$$\Theta'(t) \le \frac{\beta^2}{4}\Big(\|D\|_F^2 - 2Q_N : D + \sum_{l=0}^{N} \Delta m_l \|Q_l\|_F^2\Big) = \frac{\beta^2}{4}\Big(\|D - Q_N\|_F^2 - \int m \, \mathrm{d}\|Q\|_F^2\Big). \tag{8.8}$$

Here we used that $\Delta m_0 = m_1 - s \geq 0$; this is where the condition $m_1 \geq s$ enters. The equality in (8.8) is the summation by parts

$$\|Q_N\|_F^2 - \sum_{l=0}^{N} \Delta m_l \|Q_l\|_F^2 = \sum_{l=1}^{N} m_l \|Q_l\|_F^2 - \sum_{l=1}^{N} m_l \|Q_{l-1}\|_F^2 = \sum_{j=1}^{N} m_j \big(\|Q_j\|_F^2 - \|Q_{j-1}\|_F^2\big),$$

where we used $m_{N+1} = 1$ and $Q_0 = 0$; the right side is the penalty (4.14). The symmetry of R was not used.

Step 3: the endpoints. At $t = 1$ the site fields do not appear, and $\sum_i \boldsymbol{x}_i^{\mathsf{T}} \lambda \boldsymbol{x}_i = n\lambda : D$ on Σ_D. The integrations of $\mathbf{g}_N, \dots, \mathbf{g}_1$ act on a function of ω only and leave it unchanged, and the outermost integration gives

$$\Theta(1) = \frac{1}{ns} \log \mathbb{E}\big(Z_{\ell,n}^c(D)\big)^s + \lambda : D, \qquad Z_{\ell,n}^c(D) = \sum_{\boldsymbol{x} \in \Sigma_D} \exp\Big(\beta \sum_{a=1}^{\ell} H_n^c(x^a)\Big).$$

At $t = 0$ the Hamiltonian does not appear. Replacing Σ_D by Σ_n^ℓ increases Ψ_N, and the value increases with it, because every step of the recursion is monotone. For the unconstrained sum, Ψ_N is a sum over $i \leq n$ of functions of the independent families $(\mathbf{g}_{j,i})_{j \leq N}$, and each step $m^{-1} \log \mathbb{E} e^{m(\cdot)}$ factorizes over such sums. The contribution of each site is the one-site value $\Psi(Q, m, \lambda)$ of Definition 4.14, and the outermost integration at mass s acts on a constant. This shows that $\Theta(0) \leq \Psi(Q, m, \lambda)$.

Step 4: the completion. By Lemma 3.1, part (d), $Z_{\ell,n}^c(D) = e^{\beta \ell G} Z_{\ell,n}(D)$ and

$$\frac{1}{ns} \log \mathbb{E}\big(Z_{\ell,n}^c(D)\big)^s = p_{\ell,n}^{(s)}(D) + \frac{\beta^2 \ell^2 s}{4n};$$

indeed $G \sim N(0, 1/2)$ is independent of J, and $\mathbb{E} e^{s\beta\ell G} = e^{s^2\beta^2\ell^2/4}$. Combining Steps 2–4 with $\Theta(1) \leq \Theta(0) + \int_0^1 \Theta'(t)\, \mathrm{d}t$, we obtain (8.5). The function Θ is continuous on $[0,1]$ and differentiable on $(0,1)$, since all Gaussian integrals are finite-dimensional and U_t is Lipschitz in the Gaussian variables. Singular increments ΔQ_j cause no difficulty, since only $\Delta Q_j^{1/2}$ is used. Finally, the left side of (8.5) does not depend on the path, and the inequality passes to limits along which the right side converges. □

Remark 8.4. Gaussian interpolation for the free energy is due to Guerra [59]. Constrained comparisons for several coupled replicas go back to Talagrand [105], and the vector-spin formulation is due to Panchenko [89]; see the discussion at the start of [77, Section 2]. The quenched arbitrary-endpoint bound used in this paper is Proposition 4.19. Positive fractional moments of the SK partition function were studied by Talagrand [106]. For fixed $0 < s < 1$, the limiting normalized fractional moment is given by the one-level case of the multiscale formula of Contucci and Mingione [43, Theorem 2.1], and by the

constrained Parisi formula of Chen, Guionnet, Ko, Lacroix-A-Chez-Toine and Mourrat [35, Theorem 2.1], in which the order of the moment is a lower bound on the mass at zero. The condition $m_1 \geq s$ in Proposition 8.3 is the finite-volume form of this constraint. Quantitative finite-size versions for the one-replica partition function, with s depending on n, are due to Aronow and Lopatto [8, Section 2.3]. The replica-matrix version with a free endpoint in Proposition 8.3 is a direct adaptation of these arguments and is not stated in [77]. The completion correction $\beta^2\ell^2 s/(4n)$ is nonnegative, and below we discard it.

8.3 The profile $\max\{\alpha, 3s\}$

For $s > 0$ we define $\alpha_s(t) = \max\{\alpha(t), 3s\}$ for $0 \leq t \leq 1$. The factor three ensures that a block of three coordinates sharing a field, which is integrated at mass $\alpha_s/3$ (Lemma 4.20(b)), has mass at least s.

Lemma 8.5 (Cost of the profile)**.** *There exist $s_0 \in (0, 1/3]$ and $C_{\mathrm{fl}} < \infty$, depending only on β, such that for every $0 < s \leq s_0$ the function α_s is a distribution function and*

$$0 \leq \mathcal{P}_\beta(\alpha_s) - P_\beta \leq C_{\mathrm{fl}} s^5, \qquad \alpha_s = \alpha \text{ on } [C_{\mathrm{fl}} s, 1], \qquad \int_0^1 |\alpha_s(t) - \alpha(t)|\,\mathrm{d}t \leq C_{\mathrm{fl}} s^2. \tag{8.9}$$

Proof. Let T_*, c_1, s_0', C be the constants of Proposition 4.12, and let $s_0 = \min\{s_0', c_1 T_*/3\}$. For $0 < s \leq s_0$, part (ii) of Proposition 4.12 states that α_s is a distribution function whose cost is at most Cs^5; this is [8, Lemma 7.2, (7.3)] for this particular profile, in the fixed-temperature normalization of [8, Proposition 7.12]. Since $3s \leq c_1 T_*$, Remark 4.13 implies that $\alpha_s = \alpha$ on $[3s/c_1, 1]$ and that the L^1 distance is at most $9s^2/c_1$. The lemma holds with $C_{\mathrm{fl}} = \max\{C, 3/c_1, 9/c_1\}$. □

The cost bound concerns the explicit profile α_s, and not only the infimum of $\mathcal{P}_\beta$ over profiles with mass at least $3s$ at the origin. We need this distinction below, because we use the same profile in the comparison path for three replicas.

8.4 A change of profile before the local window

The local covariance perturbation of Section 7 starts at a time $q_0 > 0$, and on $[0, q_0]$ the three coordinates share their field, integrated at mass $\alpha/3$. For s sufficiently small that $C_{\mathrm{fl}} s < q_0$, replacing α by α_s changes only this initial segment, by Lemma 8.5. We now control the effect of that change on the perturbation.

Fix $q_0 \in (0, 1]$. For a nondecreasing $\gamma\colon [0, q_0] \to [0, 1]$ and a function $F\colon \mathbb{R} \to \mathbb{R}$ with bounded continuous derivatives of orders one to three, let Φ_F^γ be the solution on $[0, q_0] \times \mathbb{R}$ of

$$-\partial_t \Phi_F^\gamma = \frac{\beta^2}{2}\Big(\partial_{zz}\Phi_F^\gamma + \frac{\gamma(t)}{3}(\partial_z \Phi_F^\gamma)^2\Big), \qquad \Phi_F^\gamma(q_0, \cdot) = F, \tag{8.10}$$

and write $\mathcal{H}_\gamma(F) = \Phi_F^\gamma(0,0)$. For a step function γ, we interpret the equation as the recursion $\Phi_F^\gamma(t,\cdot) = T_{m,\beta^2(t_2-t)}\Phi_F^\gamma(t_2,\cdot)$ for $t_1 \le t \le t_2$, on every interval $[t_1,t_2)$ where $\gamma = 3m$. This is the recursion of three coordinates that share every field increment on $[0,q_0]$: on the line spanned by $(1,1,1)$, such an increment acts on a function of the common value of the three fields (Lemma 4.2(e)). For general γ, Φ_F^γ is the limit along step functions converging to γ in $L^1([0,q_0])$; the opening paragraph of the proof of Lemma 8.6, using the bound of Step 2 there, shows that this limit exists and does not depend on the approximating sequence. If γ is the restriction to $[0,q_0]$ of a distribution function, also denoted γ, then

$$\mathcal{H}_\gamma\big(3\Phi_\gamma(q_0,\cdot)\big) = 3\Phi_\gamma(0,0). \tag{8.11}$$

For step functions this is Lemma 4.2(d) with $k = 3$, applied at each step as in the proof of Lemma 4.20(b). The general case follows by approximation. Here we use Proposition 4.3(ii) for Φ_γ, part (a) of the next lemma, and the bound $|\mathcal{H}_\gamma(F_1) - \mathcal{H}_\gamma(F_2)| \le \|F_1 - F_2\|_\infty$, which holds by Lemma 4.2(a).

Lemma 8.6 (Response to a change of the initial profile)**.** *There exists $C_{\mathrm{r}} < \infty$, depending only on β, with the following property. Let $\gamma, \widetilde{\gamma}\colon [0,q_0] \to [0,1]$ be nondecreasing, and let $\varphi, V\colon \mathbb{R} \to \mathbb{R}$ have bounded continuous derivatives of orders one to three, with*

$$\|\varphi'\|_\infty \le 3, \qquad \|\varphi''\|_\infty \le 3, \qquad \|V'\|_\infty + \|V''\|_\infty \le 1.$$

Then:

(a) *for $F \in \{\varphi, \varphi + V\}$, $|\mathcal{H}_\gamma(F) - \mathcal{H}_{\widetilde{\gamma}}(F)| \le C_{\mathrm{r}}\|\gamma - \widetilde{\gamma}\|_{L^1([0,q_0])}$;*

(b) *the response of $\mathcal{H}$ to the terminal perturbation V changes by at most*

$$\Big|\big[\mathcal{H}_{\widetilde{\gamma}}(\varphi+V) - \mathcal{H}_{\widetilde{\gamma}}(\varphi)\big] - \big[\mathcal{H}_\gamma(\varphi+V) - \mathcal{H}_\gamma(\varphi)\big]\Big| \le C_{\mathrm{r}}\|V'\|_\infty\|\gamma - \widetilde{\gamma}\|_{L^1([0,q_0])}. \tag{8.12}$$

The quantity bounded in (8.12) is a mixed difference: it vanishes when $V = 0$ and when $\gamma = \widetilde{\gamma}$. A bound on the full difference $\mathcal{H}_{\widetilde{\gamma}}(\varphi+V) - \mathcal{H}_\gamma(\varphi+V)$ would be of order $\|\gamma - \widetilde{\gamma}\|_{L^1}$ without the factor $\|V'\|_\infty$. In the application, that full difference is accounted for by the scalar cost (8.9). The proof adapts the scalar profile-stability argument of [62, Section 2.4, Lemma 14]: subtract the two Parisi equations and represent the resulting linear equation by a diffusion. Here we also compare terminal functions, retaining the factor $\|V'\|_\infty$ in the mixed difference; this refinement is proved below.

Proof. It suffices to prove (a) and (b) for step functions $\gamma, \widetilde{\gamma}$, with constants independent of the number of steps. Indeed, let γ be general and let $\gamma_k \to \gamma$ in L^1 be step functions. The bound of Step 2 below, applied to the whole functions, shows that $\Phi_F^{\gamma_k}(t,\cdot)$ is a Cauchy sequence in the supremum norm for each t. The limit exists, it does not depend on the sequence, and the bounds pass to it. We take a common refinement, so that γ and $\widetilde{\gamma}$ are constant on the same intervals. On each such interval the solutions are smooth

in z, with bounded derivatives of orders one to three, and solve (8.10) classically; at the finitely many endpoints they are continuous in t.

We use the following maximum principle, which follows from the Feynman–Kac representation [64, Theorem 5.7.6] for a diffusion with generator $\frac{\beta^2}{2}\partial_{zz} + a\,\partial_z$, after an exponential rescaling to make the killing rate nonnegative. Let $a, c, \mathfrak{s}$ be bounded continuous functions on $[t_1, t_2] \times \mathbb{R}$, with a Lipschitz in z, and let f be bounded and solve $-\partial_t f = \frac{\beta^2}{2} f_{zz} + a f_z + c f + \mathfrak{s}$. Then

$$\|f(t,\cdot)\|_\infty \le e^{\|c\|_\infty (t_2 - t)} \Big(\|f(t_2,\cdot)\|_\infty + \int_t^{t_2} \|\mathfrak{s}(t',\cdot)\|_\infty \,\mathrm{d}t' \Big) \qquad (t_1 \le t \le t_2). \tag{8.13}$$

Applying (8.13) on consecutive intervals, we see that it holds on $[0, q_0]$ for coefficients that are continuous on each interval where γ and $\widetilde{\gamma}$ are constant.

The functions to which we apply (8.13) are bounded. Indeed, fix a step profile, and let F be L-Lipschitz. Each integration step $T_{m,v}$ maps L-Lipschitz functions to L-Lipschitz functions, by Lemma 4.2(a), and for $0 \le m \le 1$ it satisfies $\|T_{m,v} f - f\|_\infty \le \log \mathbb{E} e^{L\sqrt{v}|\zeta|}$ for L-Lipschitz f, where ζ is a standard Gaussian variable; the upper bound follows from $f(z + \sqrt{v}\,\zeta) - f(z) \le L\sqrt{v}|\zeta|$ and the monotonicity of $m \mapsto m^{-1} \log \mathbb{E} e^{mX}$, and the lower bound from Jensen's inequality. Summing over the finitely many steps, we find that $\Phi_F^\gamma(t,\cdot) - F$ is bounded. The differences Δ and $\widetilde{\mathcal{D}} - \mathcal{D}$ of Steps 2 and 3 below are bounded, and so are the derivatives to which we apply (8.13), by Step 1. The resulting estimates do not depend on the number of steps.

Step 1: derivative bounds. We use the tilted-chain martingale argument of [77, proof of Lemma 3.1], with terminal function F in place of $\log(2\cosh z)$. Let $F \in \{\varphi, \varphi + V\}$, so that $\|F'\|_\infty \le 4$ and $\|F''\|_\infty \le 4$. We claim that

$$|\partial_z \Phi_F^\gamma| \le 4, \qquad |\partial_{zz} \Phi_F^\gamma| \le 20 \qquad \text{on } [0, q_0] \times \mathbb{R}. \tag{8.14}$$

For one step $\psi = T_{m,v} f$ with $m \in [0,1]$, we have $\psi'(z) = \mathbb{E}^W f'(z + \sqrt{v}\,\zeta)$ and $\psi''(z) = \mathbb{E}^W f''(z + \sqrt{v}\,\zeta) + m \operatorname{Var}^W(f'(z + \sqrt{v}\,\zeta))$, where $\mathbb{E}^W$ and Var^W refer to the law of the standard Gaussian ζ tilted by $e^{mf(z+\sqrt{v}\,\zeta)}$; this follows by differentiating under the integral. Iterating over the steps, as in Step 1 of the proof of Lemma 8.2, we find that $\partial_z \Phi_F^\gamma(t,z)$ is the expectation of F' at the final time under the tilted chain started from z at time t, and that $\partial_{zz} \Phi_F^\gamma(t,z)$ is the expectation of F'' at the final time plus $\sum_i m_i \mathbb{E} \operatorname{Var}(M_{i+1} \mid \text{steps } 1, \dots, i)$, where M_i is the conditional expectation of the final value of F' given the first i steps. The masses $m_i = \gamma/3$ lie in $[0,1]$, and the sum of the expected conditional variances is at most the variance of the final value of F', by orthogonality of martingale increments. We conclude that $-\|F''\|_\infty \le \partial_{zz} \Phi_F^\gamma \le \|F''\|_\infty + \|F'\|_\infty^2$, which implies (8.14). These bounds do not depend on the number of steps.

Step 2: part (a). Let $\Delta = \Phi_F^\gamma - \Phi_F^{\widetilde\gamma}$. Subtracting the two equations (8.10), we have $\Delta(q_0,\cdot) = 0$ and

$$-\partial_t\Delta = \frac{\beta^2}{2}\Delta_{zz} + \frac{\beta^2\gamma}{6}\big(\partial_z\Phi_F^\gamma + \partial_z\Phi_F^{\widetilde\gamma}\big)\Delta_z + \frac{\beta^2}{6}(\gamma-\widetilde\gamma)\big(\partial_z\Phi_F^{\widetilde\gamma}\big)^2.$$

Then (8.13) with $c=0$ and (8.14) imply that $\|\Delta(t,\cdot)\|_\infty \le \frac{8\beta^2}{3}\int_t^{q_0}|\gamma-\widetilde\gamma|$, which proves (a).

We also need the analogous bound for the first derivatives. Let $\Delta_z = \partial_z\Delta = \partial_z\Phi_F^\gamma - \partial_z\Phi_F^{\widetilde\gamma}$. Differentiating (8.10) in z and subtracting, we find $\Delta_z(q_0,\cdot)=0$ and

$$-\partial_t\Delta_z = \frac{\beta^2}{2}(\Delta_z)_{zz} + \frac{\beta^2\gamma}{3}\,\partial_z\Phi_F^\gamma\,(\Delta_z)_z + \frac{\beta^2\gamma}{3}\,\partial_{zz}\Phi_F^{\widetilde\gamma}\,\Delta_z + \frac{\beta^2}{3}(\gamma-\widetilde\gamma)\,\partial_z\Phi_F^{\widetilde\gamma}\,\partial_{zz}\Phi_F^{\widetilde\gamma}.$$

By (8.14), the coefficient of Δ_z is bounded by $20\beta^2/3$ and the last term by $\frac{80\beta^2}{3}|\gamma-\widetilde\gamma|$. By (8.13) and $q_0\le 1$, we have

$$\|\Delta_z(t,\cdot)\|_\infty \le \frac{80\beta^2}{3}e^{7\beta^2}\|\gamma-\widetilde\gamma\|_{L^1([0,q_0])} \qquad (0\le t\le q_0). \tag{8.15}$$

Step 3: part (b). We write $\Phi^0=\Phi_\varphi^\gamma$, $\Phi^1=\Phi_{\varphi+V}^\gamma$, $\widetilde\Phi^0=\Phi_\varphi^{\widetilde\gamma}$ and $\widetilde\Phi^1=\Phi_{\varphi+V}^{\widetilde\gamma}$, and let $\mathcal{D}=\Phi^1-\Phi^0$ and $\widetilde{\mathcal{D}}=\widetilde\Phi^1-\widetilde\Phi^0$. The left side of (8.12) is $|\widetilde{\mathcal{D}}(0,0)-\mathcal{D}(0,0)|$. Using $(\Phi_z^1)^2-(\Phi_z^0)^2=(\Phi_z^1+\Phi_z^0)\mathcal{D}_z$, we find that $\mathcal{D}$ solves the linear equation

$$-\partial_t\mathcal{D} = \frac{\beta^2}{2}\mathcal{D}_{zz} + a\,\mathcal{D}_z, \qquad a=\frac{\beta^2\gamma}{6}\big(\Phi_z^1+\Phi_z^0\big), \qquad \mathcal{D}(q_0,\cdot)=V,$$

and $\widetilde{\mathcal{D}}$ solves the same equation with $\widetilde a = \frac{\beta^2\widetilde\gamma}{6}(\widetilde\Phi_z^1+\widetilde\Phi_z^0)$. Differentiating in z, the function $\mathcal{D}_z$ solves $-\partial_t\mathcal{D}_z = \frac{\beta^2}{2}(\mathcal{D}_z)_{zz} + a(\mathcal{D}_z)_z + a_z\mathcal{D}_z$ with $\mathcal{D}_z(q_0,\cdot)=V'$. By (8.14), we have $|a_z|\le\frac{\beta^2}{6}\cdot 40\le 7\beta^2$, and (8.13) implies that $\|\mathcal{D}_z(t,\cdot)\|_\infty\le e^{7\beta^2}\|V'\|_\infty$. Next, the difference $\widetilde{\mathcal{D}}-\mathcal{D}$ vanishes at q_0 and solves

$$-\partial_t(\widetilde{\mathcal{D}}-\mathcal{D}) = \frac{\beta^2}{2}(\widetilde{\mathcal{D}}-\mathcal{D})_{zz} + \widetilde a\,(\widetilde{\mathcal{D}}-\mathcal{D})_z + (\widetilde a - a)\mathcal{D}_z.$$

By (8.13) with $c=0$, we have

$$|\widetilde{\mathcal{D}}(0,0)-\mathcal{D}(0,0)| \le e^{7\beta^2}\|V'\|_\infty\int_0^{q_0}\|\widetilde a(t,\cdot)-a(t,\cdot)\|_\infty\,\mathrm{d}t.$$

Finally, $\widetilde a - a = \frac{\beta^2}{6}(\widetilde\gamma-\gamma)(\widetilde\Phi_z^1+\widetilde\Phi_z^0) + \frac{\beta^2\gamma}{6}\big((\widetilde\Phi_z^1-\Phi_z^1)+(\widetilde\Phi_z^0-\Phi_z^0)\big)$. Using (8.14) for the first term and (8.15) for the second, we find $\int_0^{q_0}\|\widetilde a-a\|_\infty\,\mathrm{d}t \le C\|\gamma-\widetilde\gamma\|_{L^1([0,q_0])}$ with C depending only on β. This proves (8.12). □

8.5 The local fractional bound

We now combine Proposition 8.3, Lemma 8.5 and Lemma 8.6 with the cubic deficit of Section 7. We first recall what we use from that section. Fix the ratio $\vartheta \in (0, 1/2]$ and the amplitude $E > 0$ of Section 7.6, satisfying the two strict inequalities (7.37); in the dynamical argument, $\vartheta = 1/2$ and $E = 1/100$. Let $G(\vartheta, E) > 0$ be the corner coefficient of Section 7.6. Fix compact intervals $\mathcal{J} \subset \operatorname{int} I$ with $I \subset (0, q_\beta)$, and let $m_\beta > 0$ be a lower bound on I for the scalar factor $M(r) = \alpha(r)^2 \beta^6 \mathbb{E}\chi(r, Y_r)^3$ of Proposition 7.13, as in (5.20). For $D = D(r, q, \delta)$ with $d = q - r > 0$, we use the parameters of (7.1), namely $h = d/\vartheta$, $q_0 = r - h$ and $k = \delta/h$. Let $\mathcal{T}(q_0, h, k)$ be the trial value (7.8), let $V = V(q_0, h, k, \cdot)$ be the boundary perturbation (7.11), and let $\Pi = \Pi(q_0, h, k)$ be the penalty of Lemma 7.3. We use the following three facts.

(T1) By (7.8) and Lemma 7.3, $\mathcal{T}(q_0, h, k)$ is the limit of the trial values of admissible paths $(Q^{(N)}, m^{(N)})$ with endpoint D and multiplier zero. On $[0, q_0]$ these paths have the increments $\mathbf{e}\mathbf{e}^{\mathsf{T}}\,\mathrm{d}t$ at the masses $\alpha/3$, evaluated at the left endpoints of a partition; on $[q_0, 1]$ their masses are at least $\alpha(q_0)/3$; and all masses are nondecreasing. Moreover

$$\mathcal{T}(q_0, h, k) = \mathcal{H}_\alpha\big(3\Phi(q_0, \cdot) + V\big) - \frac{\beta^2}{4}\Pi,$$

where $\mathcal{H}_\alpha$ is the operator of Section 8.4 on $[0, q_0]$ with $\gamma = \alpha$. The function V and the penalty $\Pi_+ = \Pi - 6\int_0^{q_0} t\alpha(t)\,\mathrm{d}t$ of the path on $[q_0, 1]$ depend on α only through its restriction to $[q_0, 1]$.

(T2) By the spatial regularity in Lemma 7.4, V has bounded continuous derivatives of orders one to three. By (7.14), $\|V\|_{C_b^2} \le C_V h^2 k^2 = C_V \delta^2$ for $h \le h_0$. We use only the consequence $\|V\|_{C_b^2} \le C_V \delta^2/h$ for $h \le \min\{h_0, 1\}$; no smallness of the third derivative is needed.

(T3) By Proposition 7.15, applied to the interval $\mathcal{J}$ of centers (note that $m_\beta \le \inf_{\mathcal{J}} M$), for every $\bar\sigma \in (0, G(\vartheta, E))$ there exists $d_0 > 0$ such that $\mathcal{T}(q_0, h, k) \le 3P_\beta - m_\beta E^2\{G(\vartheta, E) - \bar\sigma\}\rho^3$, which is (7.38), for every $0 < \rho \le d_0$ and every feasible $D(r, q, \delta)$ with $r \in \mathcal{J}$, $\rho \le q - r \le d_0$ and $E\rho \le |\delta| \le 2E\rho$; for these matrices also $h \le h_0$. The window is $[q_0, T]$ with $T = q_0 + h\tau_T$ and $\tau_T = 2 + 4k^2 L_\vartheta$, by (7.3). If $|\delta| \le 2Ed$, then $|k| \le 2\vartheta E$, and the window has length $T - q_0 \le C_{\mathrm{win}} d$ with $C_{\mathrm{win}} = (2 + 16\vartheta^2 E^2 L_\vartheta)/\vartheta$.

By (8.11), $\mathcal{H}_\alpha(3\Phi(q_0, \cdot)) = 3\Phi(0, 0)$.

Now let $\gamma\colon [0, 1] \to [0, 1]$ be nondecreasing, with $\gamma = \alpha$ on $[q_0, 1]$ and $\gamma \le \alpha(q_0)$ on $[0, q_0]$. We write $\mathcal{T}_\gamma(q_0, h, k)$ for the limit of the trial values of the paths of (T1) with the masses $\alpha/3$ on $[0, q_0]$ replaced by the left-endpoint values of $\gamma/3$. These paths are admissible, with the same endpoint, and on $[q_0, 1]$ nothing changes. The penalty of the segment $[0, q_0]$ converges to $\int_0^{q_0} \frac{\gamma}{3}\,\mathrm{d}(9t^2) = 6\int_0^{q_0} t\gamma(t)\,\mathrm{d}t$. By Lemma 8.6(a) and the bound

$|\mathcal{H}_\gamma(F_1) - \mathcal{H}_\gamma(F_2)| \le \|F_1 - F_2\|_\infty$, the values converge as well, and with $\varphi = 3\Phi(q_0, \cdot)$ we find

$$\mathcal{T}_\gamma(q_0, h, k) = \mathcal{H}_\gamma(\varphi + V) - \frac{3\beta^2}{2}\int_0^{q_0} t\gamma(t)\,\mathrm{d}t - \frac{\beta^2}{4}\Pi_+. \tag{8.16}$$

For $\gamma = \alpha$, this is the identity of (T1). If γ is a distribution function, then by (8.11) we have

$$3\mathcal{P}_\beta(\gamma) = \mathcal{H}_\gamma(\varphi) - \frac{3\beta^2}{2}\int_0^{q_0} t\gamma(t)\,\mathrm{d}t - \frac{3\beta^2}{2}\int_{q_0}^{1} t\alpha(t)\,\mathrm{d}t. \tag{8.17}$$

Subtracting (8.16) and (8.17) for α from the same identities for γ, and using $\mathcal{P}_\beta(\alpha) = P_\beta$, we obtain

$$\begin{aligned}\mathcal{T}_\gamma(q_0, h, k) - \mathcal{T}(q_0, h, k) = 3\big(\mathcal{P}_\beta(\gamma) - P_\beta\big)\\ + \big[\mathcal{H}_\gamma(\varphi + V) - \mathcal{H}_\gamma(\varphi)\big] - \big[\mathcal{H}_\alpha(\varphi + V) - \mathcal{H}_\alpha(\varphi)\big].\end{aligned} \tag{8.18}$$

Lemma 8.7 (Local fractional bound)**.** *Let $\mathcal{J} \subset \operatorname{int} I$ and $I \subset (0, q_\beta)$ be compact intervals, let ϑ, E, m_β be as above, and let $0 < g < m_\beta E^2 G(\vartheta, E)$. There exist $d_0, S, \rho_0 > 0$, depending only on β, $\mathcal{J}$, I, ϑ, E and g, such that the following holds. For every $0 < \rho \le \rho_0$, every $n \ge 2$ and every feasible $D = D(r, q, \delta)$ with*

$$r \in \mathcal{J}, \qquad \rho \le q - r \le d_0, \qquad E\rho \le |\delta| \le 2E\rho,$$

the window $[q_0, T]$ of the local perturbation lies in I, and with $s = S\rho$ we have

$$\frac{1}{ns}\log \mathbb{E}Z_{3,n}(D)^s \le 3P_\beta - g\rho^3. \tag{8.19}$$

The constants d_0, S and ρ_0 do not depend on n, on ρ, or on (r, q, δ) in the stated domain. The order in which they are chosen is recorded in Table 8.1.

Parameter	Chosen in terms of	Purpose
$\mathcal{J}$, I, ϑ, E, m_β	fixed first	centers; geometry of Section 7
$g < g'$	$m_\beta E^2 G(\vartheta, E)$	two strict margins (8.20)
d_0	g', $\mathcal{J}$, I, ϑ, E	cubic deficit (T3); window in I; $h \le 1$
S	g, g', ϑ, E	profile response at most $\frac{1}{2}(g' - g)\rho^3$
ρ_0	all of the above	α_s admissible; V small; cost s^5
$n \ge 2$	arbitrary	the bound is uniform in n

Table 8.1. Order of the choices in the proof of Lemma 8.7.

Proof. The proof has three steps. We first fix the constants. We then apply Proposition 8.3 to the comparison path with the profile α_s on the common segment, and finally we compare its trial value with $\mathcal{T}(q_0, h, k)$ by (8.18).

Step 1: choice of constants. Fix g' with

$$0 < g < g' < m_\beta E^2 G(\vartheta, E). \tag{8.20}$$

Let d_0 be the constant of (T3) for $\bar\sigma = G(\vartheta,E) - g'/(m_\beta E^2)$, decreased so that also $d_0 \le \vartheta$ and $(1/\vartheta + C_{\mathrm{win}})d_0 \le \operatorname{dist}(\mathcal{J}, \mathbb{R}\setminus I)$. Then, for every $0 < \rho \le d_0$ and every D in the domain of the lemma,

$$\mathcal{T}(q_0,h,k) \le 3P_\beta - g'\rho^3, \tag{8.21}$$

moreover $h = d/\vartheta \le 1$, and the window $[q_0,T] \subset [r - d/\vartheta, r - d/\vartheta + C_{\mathrm{win}}d]$ lies in I, so that $q_0 \ge a_*$ with $a_* = \min I > 0$. Let C_{fl}, s_0 and C_{r} be the constants of Lemmas 8.5 and 8.6, let $C' = 3C_{\mathrm{fl}}$ and $C'' = C_{\mathrm{r}}C_V C_{\mathrm{fl}}$, and choose $S \in (0, 1/3]$ so small that

$$4\vartheta E^2 C'' S^2 \le \tfrac12 (g' - g). \tag{8.22}$$

Finally choose $\rho_0 \in (0, d_0]$ so small that

$$S\rho_0 \le s_0, \qquad C_{\mathrm{fl}} S\rho_0 < a_*, \qquad 8\vartheta E^2 C_V \rho_0 \le 1, \qquad C' S^5 \rho_0^2 \le \tfrac12 (g' - g).$$

Now fix $0 < \rho \le \rho_0$, $n \ge 2$ and D in the domain of the lemma, and set $s = S\rho$.

Step 2: the fractional comparison. Since $s \le s_0$ and $C_{\mathrm{fl}} s < a_* \le q_0$, Lemma 8.5 shows that α_s is a distribution function with $\alpha_s = \alpha$ on $[q_0, 1]$. Moreover $\alpha_s \le \alpha_s(q_0) = \alpha(q_0)$ on $[0, q_0]$, and $\mathcal{T}_{\alpha_s}(q_0,h,k)$ is defined. By its definition, it is the limit of the trial values of admissible paths with endpoint D and multiplier zero, whose masses are left-endpoint values of $\alpha_s/3$ on $[0,q_0]$ and are at least $\alpha(q_0)/3$ on $[q_0,1]$. All these masses are at least s, because $\alpha_s \ge 3s$ and $\alpha(q_0) = \alpha_s(q_0) \ge 3s$. By Proposition 8.3 with $\ell = 3$ for these paths, we obtain

$$p^{(s)}_{3,n}(D) + \frac{9\beta^2 s}{4n} \le \mathcal{T}_{\alpha_s}(q_0,h,k). \tag{8.23}$$

Step 3: comparison with $\mathcal{T}(q_0,h,k)$. We apply (8.18) with $\gamma = \alpha_s$. The first term on its right side is at most $3C_{\mathrm{fl}}s^5 = C's^5$ by (8.9). For the second, note that $\varphi = 3\Phi(q_0,\cdot)$ satisfies $\|\varphi'\|_\infty \le 3$ and $\|\varphi''\|_\infty \le 3$, since $|u| \le 1$ and $0 < \chi \le 1$ by Proposition 4.3(i), and that $\|V'\|_\infty + \|V''\|_\infty \le 2C_V\delta^2/h \le 8\vartheta E^2 C_V \rho \le 1$ by (T2), since $|\delta| \le 2E\rho$ and $h = d/\vartheta \ge \rho/\vartheta$. Lemma 8.6(b) on $[0,q_0]$, with $\widetilde\gamma = \alpha_s$ and $\gamma = \alpha$, and (8.9) then bound the mixed difference by $C_{\mathrm{r}}\|V'\|_\infty \|\alpha_s - \alpha\|_{L^1} \le C_{\mathrm{r}}C_V(\delta^2/h)C_{\mathrm{fl}}s^2$. We have shown that the change of profile costs at most

$$\mathcal{T}_{\alpha_s}(q_0,h,k) - \mathcal{T}(q_0,h,k) \le C's^5 + C''\frac{\delta^2 s^2}{h}. \tag{8.24}$$

Since $|\delta| \le 2E\rho$, $h = d/\vartheta$ and $d \ge \rho$, we have

$$\frac{\delta^2 s^2}{h} = \frac{\vartheta\,\delta^2 S^2 \rho^2}{d} \le 4\vartheta E^2 S^2 \rho^3.$$

Inserting (8.21) and (8.24) into (8.23), and discarding the nonnegative term $9\beta^2 s/(4n)$, we obtain

$$p_{3,n}^{(s)}(D) \le 3P_\beta - g'\rho^3 + C'S^5\rho^5 + 4\vartheta E^2 C'' S^2 \rho^3 \le 3P_\beta - g\rho^3,$$

where in the last step we used (8.22) and $C'S^5\rho^2 \le \frac{1}{2}(g' - g)$. This is (8.19). □

Remark 8.8. The bound (8.19) is a deterministic statement about the Gaussian moments $\mathbb{E}Z_{3,n}(D)^s$. It holds for every $n \ge 2$ once $\rho \le \rho_0$, and it requires no relation between ρ and n. The separation $q_0 \ge a_* > 0$ between the local window and the origin is what makes the change of profile harmless: since $\alpha_s = \alpha$ on $[q_0, 1]$, the local window, its terminal correction and its penalty are identical for the two profiles. The scalar cost of order s^5 is much smaller than needed; a cost of order s^3 would also suffice, after decreasing S. Since $\|V\|_{C_b^2} \le C_V\delta^2$ by (7.14), the response term is in fact of order $\delta^2 s^2 \le 4E^2S^2\rho^4$; we used only the weaker bound $C_V\delta^2/h$, for which the small choice of S is needed. The local bound concerns only $q - r \le d_0$. For $q - r > d_0$ the construction of Section 7 is not used, and Section 9 uses instead the quenched comparison of Lemma 9.5, whose deficit is a multiple of δ^2, much larger than ρ^3.

SECTION 9

A simultaneous estimate for overlap constraints

In this section we prove Lemma 9.7. For deterministic scales $\rho_n \to 0$ and η_n satisfying the conditions (S1)–(S5) below, it provides a disorder event $\mathcal{A}_n$ of probability tending to one on which three families of overlap configurations have small Gibbs probability simultaneously: triples whose overlaps lie in a narrow band around a center r, triples violating a small-overlap constraint across a large overlap, and pairs whose absolute overlap $q = |R(x, y)|$ has $|q - \Gamma(q)|$ bounded away from zero. These estimates are the static input of the dynamical argument: Section 10 uses them to choose families of reference configurations (Lemma 10.2), and Sections 11 and 12 use them along trajectories of the heat-bath dynamics. The conditions (S1)–(S5) hold at the scales of Theorem 1.1, and more generally under (10.10) and (10.11); see Remark 9.8.

The narrow triples split into two regimes. When the larger overlap q is close to the center r, we use the local fractional bound Lemma 8.7, which rests on the cubic deficit of Section 7. When $q - r$ is bounded below, we use a quadratic pressure deficit, which we derive in Lemma 9.5 from the three-replica comparison of [77, Section 7], with a fixed width and a fixed tolerance. The pair bound and the fixed-scale bound are the pressure estimates Propositions 4.23 and 4.24, which are [77, Lemma 4.3] and [77, (4.20)]. In all these regimes the mechanisms are those of [77]; what changes are the scales at which the pressure deficits are converted into probability bounds. The simultaneous event and its probability estimate are new relative to [77]: they replace the conversion [77, Lemma 4.2], whose finite-size error $n^{-1/24}$ is far too large at the scale ρ_n^3.

Two levels of probability. We keep two kinds of error separate, and the later sections depend on this distinction. The first is the disorder probability $\mathbb{P}(\mathcal{A}_n^c)$, for which we obtain the upper bound $O(n^{-2/15})$ at the scales of Theorem 1.1; it is paid once. The second is the Gibbs probability of the forbidden configurations at a fixed disorder in $\mathcal{A}_n$, which for narrow triples is bounded above by $C(n+1)^3 e^{-cn^{3/10}}$ at the same scales. Only errors of the second kind are multiplied by the number of reference configurations and paths in Sections 10 to 12; see Table 9.1 and Remark 9.9.

Upper bound on the mean bias. The proof of Lemma 9.7 uses the mean bias $\Delta_{n,\beta} = P_\beta - p_n(\beta)$ only through the hypothesis (S1). The bound $\Delta_{n,\beta} \le C_b n^{-7/10}$ of Theorem 4.11 gives the scale $\rho_n \asymp n^{-7/30}$ of Theorem 1.1. The variance bound of Theorem 4.10 is used only for the lower bound (A1) on F_n below, through Chebyshev's inequality and the

hypothesis (S2).

Standing notation. We keep the notation of Section 8; in particular we do not abbreviate β^2. Triples are taken in the order (v,x,y), and for a triple we write

$$r=\frac{R(v,x)+R(v,y)}{2},\qquad \delta=\frac{R(v,x)-R(v,y)}{2},\qquad q=R(x,y),$$

so that its overlap matrix is the matrix $D(r,q,\delta)$ of (4.16). We also write $\chi_t=\chi(t,Y_t)$. Constants C,c depend only on β and on the fixed parameters named in each statement; they never depend on n, on the overlap parameters (r,q,δ), or on the scales ρ_n,η_n.

9.1 Orientations, pairs and the fixed scale

Every overlap takes one of the $n+1$ values $-1,-1+2/n,\dots,1$, so there are at most $n+1$ feasible matrices $D_2(\bar q)$ and at most $(n+1)^3$ feasible matrices $D(r,q,\delta)$. Let $\mathfrak{O}_\ell$ be the group of maps $\mathsf{O}\colon (x^a)_{a\le\ell}\mapsto(\epsilon_a x^{\varpi(a)})_{a\le\ell}$ of Remark 4.17, where ϖ is a permutation of $\{1,\dots,\ell\}$ and $\epsilon\in\{-1,1\}^\ell$; the group $\mathfrak{O}_3$ has 48 elements. By Remark 4.17, $Z_{\ell,n}(\mathsf{O}D)=Z_{\ell,n}(D)$ for every disorder and every orientation O. In addition, every $\mathsf{O}\in\mathfrak{O}_\ell$ is a bijection of Σ_n^ℓ that preserves the product measure $(\pi^J)^{\otimes\ell}$, because $\pi^J(-x)=\pi^J(x)$. We have

$$(\pi^J)^{\otimes\ell}(\mathsf{O}B)=(\pi^J)^{\otimes\ell}(B)\qquad(B\subset\Sigma_n^\ell,\ \mathsf{O}\in\mathfrak{O}_\ell).\tag{9.1}$$

All statements below are proved for canonical orientations. A disorder bound for $Z_{\ell,n}(D)$ holds verbatim for $Z_{\ell,n}(\mathsf{O}D)$, a Gibbs bound for a set B holds for $\mathsf{O}B$, and a union over all orientations of a triple costs at most a factor 48.

Lemma 9.1 (Pair constraints)**.** *For every $n\ge 2$ and every feasible $D_2(\bar q)$,*

$$p_{2,n}\big(D_2(\bar q)\big)\le 2P_\beta-\frac12\big(|\bar q|-\Gamma(|\bar q|)\big)^2.\tag{9.2}$$

Moreover, there exist $q_{\mathrm{tail}}<1$ and $c_{\mathrm{tail}}>0$, depending only on β, such that $q-\Gamma(q)\ge c_{\mathrm{tail}}$ for $q_{\mathrm{tail}}\le q\le 1$.

Proof. For $\bar q\ge 0$, (9.2) is Proposition 4.23, valid for every n and every feasible pair constraint; for $\bar q<0$ it follows from $Z_{2,n}(D_2(\bar q))=Z_{2,n}(D_2(|\bar q|))$. By Proposition 4.3(iv), Γ is continuous, and $\Gamma(1)=\mathbb{E}\tanh^2(Y_1)<1$ because $u(1,z)=\tanh z$. Since $q-\Gamma(q)$ is continuous and positive at $q=1$, there exist q_{tail} and c_{tail} as claimed. □

Lemma 9.2 (The fixed scale)**.** *For every $b_*\in(0,1)$ there exist $\underline{b}(b_*),\nu_{\mathrm{fix}}(b_*),c_{\mathrm{fix}}(b_*)>0$, depending only on β and b_*, such that*

$$p_{3,n}\big(D(r,q,\delta)\big)\le 3P_\beta-c_{\mathrm{fix}}(b_*)\,\delta^2\tag{9.3}$$

for every $n \geq 2$ and every feasible $D(r, q, \delta)$ with $b_* \leq q \leq 1$, $|r + \delta| \leq \underline{b}(b_*)$, $|r - \delta| \leq \underline{b}(b_*)$ and $|q - \Gamma(q)| \leq \nu_{\mathrm{fix}}(b_*)$. The same holds after decreasing $\underline{b}(b_*)$ or $\nu_{\mathrm{fix}}(b_*)$.

Proof. Let r_0, δ_0, ν, c_0 be the constants of Proposition 4.24 for the cutoff b_*, and set $\underline{b}(b_*) = \min\{r_0, \delta_0\}$, $\nu_{\mathrm{fix}}(b_*) = \nu$ and $c_{\mathrm{fix}}(b_*) = c_0$. If $|r \pm \delta| \leq \underline{b}(b_*)$, then $|r| \leq \underline{b}(b_*)$ and $|\delta| \leq \underline{b}(b_*)$. For $r \geq 0$ the bound is Proposition 4.24. For $r < 0$, reversing v maps $D(r, q, \delta)$ to $D(-r, q, -\delta)$, and the two constrained partition functions are equal by Remark 4.17. Decreasing $\underline{b}(b_*)$ or $\nu_{\mathrm{fix}}(b_*)$ shrinks the set of matrices to which the bound is applied. □

Here we use only the pressure bound [77, (4.20)]. The fixed-scale probability estimate [77, Lemma 4.1], obtained using the conversion of [77, Lemma 4.2], concerns overlap imbalances of size $n^{-1/100}$; below we derive the probability bound at overlap imbalances of size η_n. No restriction $r \geq |\delta|$ is needed, so the two small overlaps $r \pm \delta$ may have opposite signs.

9.2 Separated overlaps

We now treat triples $D(r, q, \delta)$ with r in a compact subinterval of $(0, q_\beta)$ and $q - r$ bounded below. The argument is that of [77, Section 7, Steps 1–8], recalled in Proposition 4.25. There it is carried out with the shrinking parameters $h = \rho_{\mathrm{OB}}^8$ and $\theta = \rho_{\mathrm{OB}}^2$ and a restriction on the Parisi mass of $[r - h, r + h]$. We need it with fixed parameters, so we write the argument out in this form; Remark 4.26 lists the places where the scales enter. The only change in the computation is that the oscillation of α near r is kept as an explicit term instead of being bounded by ρ_{OB}^4. We begin with the scalar estimate that produces the deficit.

Lemma 9.3 (Decay of the mixed susceptibility)**.** *Let $\mathcal{J} \subset (0, q_\beta)$ be a compact interval. There exists $c_\chi > 0$, depending only on β and $\mathcal{J}$, such that for every $r \in \mathcal{J}$ and $r \leq q \leq 1$,*

$$\beta^2 \mathbb{E}[\chi_q \chi_r] = 1 - \beta^4 \int_r^q \alpha(t)\, \mathbb{E}[\chi_t^2 \chi_r]\, \mathrm{d}t \leq 1 - c_\chi (q - r). \tag{9.4}$$

Proof. By Corollary 4.8, the hypotheses of Proposition 4.25 hold for every $t_0 \in (0, q_\beta/3]$; note that then $3t_0 \leq q_\beta < 1$. The intervals $(t_0/2, 3t_0)$ with $t_0 \in (0, q_\beta/3]$ cover $(0, q_\beta)$, and finitely many of them cover the compact set $\mathcal{J}$. By part (ii) of Proposition 4.25, which is [77, (7.31)], the bound (9.4) holds for r in each of these intervals, with a constant depending on t_0, and we take c_χ to be the smallest of these constants. □

The identity in (9.4) follows from the conditional evolution of χ in Proposition 4.3(iv), multiplied by χ_r, and from $\beta^2 \mathbb{E}\chi_r^2 = \Gamma'(r) = 1$. The inequality comes from restricting the scalar diffusion to an event of positive probability on which $|Y_t|$ stays bounded, where χ has a positive minimum; no lower bound on χ on the whole real line is used. The next lemma is the fixed-parameter form of [77, (7.29)–(7.30)].

Lemma 9.4 (Comparison with a triangular covariance perturbation of fixed width)**.** *Let $r, q \in (0,1]$, $\bar h \in (0,1]$ and $\bar\theta \in (0, 1/2]$ satisfy $\bar h < r$ and $r + \bar h < q$, and set*

$$\bar W = \frac{2(1-\bar\theta)^2 + \bar\theta^2}{\bar h}, \qquad \mathrm{osc}_r = \alpha(r+\bar h) - \alpha(r-\bar h).$$

There exist $\widetilde C, \widehat C < \infty$, depending only on β, such that for every $n \geq 2$ and every feasible $D(r,q,\delta)$ with $8\beta^2 \bar W \delta^2 \leq 1$,

$$\begin{aligned} p_{3,n}\big(D(r,q,\delta)\big) \leq 3P_\beta + \delta^2 \Big\{ &\beta^2 \bar W \alpha(q)\big(q - \Gamma(q)\big) + 2(1-\bar\theta)\bar\theta\beta^2\alpha(r)\big(\beta^2 \mathbb{E}[\chi_q \chi_r] - 1\big) \\ &+ \widetilde C(\mathrm{osc}_r + \bar\theta^2) \Big\} + \widehat C\big(|\delta|^3 \bar h^{-3/2} + \delta^4 \bar h^{-2}\big). \end{aligned} \tag{9.5}$$

Proof. We perturb the base path of Lemma 4.20(b) for three replicas by a contrast of fixed width $\bar h$ around r, keep the resulting endpoint mismatch of order δ^2, and evaluate the change of the one-site value with Gaussian integration by parts. The proof follows [77, Section 7, Steps 1–7] and has seven steps. Steps 1–6 are carried out for step approximations of α, and Step 7 passes to the limit and applies (4.9).

Step 1: step data and the base path. Fix partitions of $[0,1]$ whose mesh tends to zero and which contain $r - \bar h$, r, $r + \bar h$ and q. On each interval $[t_{i-1}, t_i)$ of such a partition let the step function take the value $\alpha(t_{i-1})$, and let it equal 1 at $t = 1$. These step functions are distribution functions, they converge to α in $L^1([0,1])$, they agree with α at r and q, and on $[r - \bar h, r + \bar h]$ they take values in $[\alpha(r-\bar h), \alpha(r+\bar h)]$. In Steps 1–6 we fix one of them and denote it, and the scalar objects built from it, again by α, Φ, u, χ; its oscillation on $[r - \bar h, r + \bar h]$ is at most osc_r. The bounds $|u| \leq 1$, $0 < \chi \leq 1$ and $|\chi(t,z) - \chi(t,z')| \leq C|z - z'|$ of Proposition 4.3(i) hold uniformly over distribution functions, and all constants below are independent of the partition.

Let (Q, m) be the base path of Lemma 4.20(b) for this step function. On $[0, r]$ the three coordinates share every increment, at mass $\alpha/3$. On $[r, q]$ the coordinate v has its own increments at mass α, and the coordinates x, y share increments at mass $\alpha/2$. On $[q, 1]$ the three coordinates have independent increments at mass α. By Lemma 4.20(b), its endpoint is $D(r,q,0)$ and its trial value for $D(r,q,0)$ is $3\mathcal{P}_\beta(\alpha)$ [77, (7.7)]. Let z^x and z^v be the raw coordinates of the cumulative fields of x and of v, sums of the Gaussian increments, which agree on $[0, r]$. We write $\mathbb{E}_0$ for the path law of the base path. Under the Gaussian law of the increments, z^x and z^v are Brownian motions with variance β^2 per unit time, observed at the partition points; under $\mathbb{E}_0$, they are the paths Y^x and Y^v of Lemma 4.20(d). Under $\mathbb{E}_0$, each is a scalar diffusion for α, they coincide on $[0, r]$, and conditionally on the common path through r their continuations are independent [77, (4.6)].

Step 2: the perturbation and its endpoint. We define the Gaussian linear combinations

$$\mathsf{Y}_0 = \frac{(z^x_r - z^x_{r-\bar h}) - (z^x_{r+\bar h} - z^x_r)}{\bar h}, \qquad \mathsf{Y}_1 = \frac{z^v_{r+\bar h} - z^v_r}{\bar h}, \qquad \mathsf{Y} = (1-\bar\theta)\mathsf{Y}_0 + \bar\theta \mathsf{Y}_1,$$

which are [77, (7.8)] with $h = \bar h$ and $\theta = \bar\theta$. At every increment of the first two stages, we add δ times its contribution to Y to the field of x and subtract it from the field of y; the field of v is unchanged. Every modified increment is a linear image of the original Gaussian increment, so it is positive semidefinite, increments remain independent, and the masses are unchanged. At the end of the second stage the added fields are $(0, \delta\mathsf{Y}, -\delta\mathsf{Y})$. In units of β^2, the variance of Y is $\bar W$, its covariance with the final field of v is $(1-\bar\theta) + \bar\theta = 1$, and its covariance with the final fields of x and of y is $(1-\bar\theta) - (1-\bar\theta) = 0$. Let W_c, W_v and W_x be the corresponding accumulated quantities along the path, and let

$$L_1 = \begin{pmatrix} 0 & W_v & -W_v \\ W_v & 2W_x & 0 \\ -W_v & 0 & -2W_x \end{pmatrix}, \qquad L_2 = (\mathbf{e}_x - \mathbf{e}_y)(\mathbf{e}_x - \mathbf{e}_y)^{\mathsf T}.$$

Then the perturbed covariance is $\widetilde Q = Q + \delta L_1 + \delta^2 W_c L_2$ along the path, and its endpoint is $\widetilde Q_N = D(r,q,\delta) + \delta^2 \bar W L_2$ [77, (7.10), (7.12)]. Its diagonal is not equal to one, which the arbitrary-endpoint comparison Proposition 4.19 allows; the mismatch cost is

$$\frac{\beta^2}{4} \| D(r,q,\delta) - \widetilde Q_N \|_F^2 = \beta^2 \delta^4 \bar W^2. \tag{9.6}$$

Step 3: the penalty. Note that $Q_{vx} = Q_{vy}$ and $Q_{xx} = Q_{yy}$ along the base path, which implies that $Q : L_1 = 0$; also $L_1 : L_2 = 0$, $\|L_1\|_F^2 = 4W_v^2 + 8W_x^2$ and $\|L_2\|_F^2 = 4$. Expanding, we have

$$\|\widetilde Q\|_F^2 - \|Q\|_F^2 = \delta^2 \big(4W_v^2 + 8W_x^2 + 2W_c\, Q : L_2\big) + 4\delta^4 W_c^2.$$

In the first two stages the base fields of x and y coincide and $Q : L_2 = 0$; in the last stage, $W_c = \bar W$ and $Q : L_2 = 2(t-q)$. The last term has nonnegative increments, because W_c is nondecreasing. We conclude that the increase of the penalty is at least

$$\frac{\beta^2}{4}\Big(\int m\, \mathrm{d}\|\widetilde Q\|_F^2 - \int m\, \mathrm{d}\|Q\|_F^2 \Big) \ge \beta^2\delta^2 \Big(\bar W \int_q^1 \alpha(t)\, \mathrm{d}t + \mathcal{I} \Big), \qquad \mathcal{I} = \int m\ \mathrm{d}\big(W_v^2 + 2W_x^2\big). \tag{9.7}$$

The integral $\mathcal{I}$ has three pieces. On $[r-\bar h, r]$, the common increments carry mass $\alpha/3$ and $W_v = W_x$ grows from 0 to $1-\bar\theta$; this piece is $\int \frac{\alpha}{3}\ \mathrm{d}(3W_v^2)$. On the v-branch over $[r, r+\bar h]$, the mass is α and W_v grows from $1-\bar\theta$ to 1. On the shared branch of x, y over $[r, r+\bar h]$, the mass is $\alpha/2$ and W_x decreases from $1-\bar\theta$ to 0, and this piece is $\int \frac{\alpha}{2}\ \mathrm{d}(2W_x^2)$. After cancelling the block-size factors, the three pieces are integrals of α against $\mathrm{d}(W_v^2)$, $\mathrm{d}(W_v^2)$ and $\mathrm{d}(W_x^2)$, respectively. Their integrators are monotone, with total variations $(1-\bar\theta)^2$,

$2\bar\theta - \bar\theta^2$ and $(1-\bar\theta)^2$, each at most one, and on these pieces α differs from $\alpha(r)$ by at most osc_r. Replacing α by $\alpha(r)$ therefore changes their sum by at most $3\,\mathrm{osc}_r$:

$$\Big|\mathcal{I} - \alpha(r)\big\{(1-\bar\theta)^2 + (2\bar\theta-\bar\theta^2) - (1-\bar\theta)^2\big\}\Big| \le 3\,\mathrm{osc}_r,$$

so that $\mathcal{I} \ge \alpha(r)(2\bar\theta-\bar\theta^2) - 3\,\mathrm{osc}_r$. This is [77, (7.14)–(7.15)], with $3\,\mathrm{osc}_r$ in place of the error $O(\rho_{\mathrm{OB}}^4)$.

Step 4: the one-site value. Let Ψ_0 and $\widetilde\Psi$ be the one-site values of the base and perturbed paths. The last stage is the same for both, and integrating it leaves the sum of $\Phi(q,\cdot)$ over the three fields. With $z = z_q^x$ the common field of x, y at the end of the second stage, the two boundary functions differ by

$$\mathrm{Y} = \Phi(q, z+\delta\mathrm{Y}) + \Phi(q, z-\delta\mathrm{Y}) - 2\Phi(q,z),$$

a function of the Gaussian increments of the first two stages. We claim that

$$\widetilde\Psi - \Psi_0 \le \log \mathbb{E}_0 e^{\mathrm{Y}}. \tag{9.8}$$

Indeed, both hierarchies integrate the same Gaussian increments at the same masses. Let Δ_i be the difference of their partial values after the increments beyond i have been integrated, as a function of the increments through i, and let $\mathcal{K}_i$ be the tilted kernel of the base path at increment i. Taking the ratio of the two defining integrals, we find $\Delta_{i-1} = m^{-1}\log \mathcal{K}_i e^{m\Delta_i}$ if increment i has mass $m > 0$, and $\Delta_{i-1} = \mathcal{K}_i\Delta_i$ if $m = 0$. Jensen's inequality, together with the concavity of $x \mapsto x^m$ for $0 < m \le 1$, implies that $\Delta_{i-1} \le \log \mathcal{K}_i e^{\Delta_i}$ in both cases. If $\Delta_i \le \log \mathbb{E}_0[e^{\mathrm{Y}} \mid \text{increments through } i]$, then the tower property gives the same inequality at $i-1$. At the end of the second stage we have $\Delta = \mathrm{Y}$, and induction proves (9.8). This is the kernel comparison [77, (4.11), (7.17)]; it uses no inverse of the field transformation, and singular increments cause no difficulty.

Using convexity of $\Phi(q,\cdot)$, the bound $\chi \le 1$ and the Lipschitz bound on χ, we find by the second-order Taylor formula with integral remainder that

$$0 \le \mathrm{Y} \le \delta^2\mathrm{Y}^2, \qquad \big|\mathrm{Y} - \delta^2\chi_q\mathrm{Y}^2\big| \le C|\delta\mathrm{Y}|^3, \qquad \chi_q = \chi(q, z_q^x). \tag{9.9}$$

Under $\mathbb{E}_0$, the diffusion representation gives $\mathrm{Y} = \mathrm{Y}^{\mathrm{G}} + \mathrm{Y}^{\mathrm{B}}$, where Y^{G} is a centered Gaussian variable of variance $\varsigma^2 = \beta^2\bar W$ and $|\mathrm{Y}^{\mathrm{B}}| \le 2\beta^2$, since every drift is bounded by β^2, is integrated over an interval of length $\bar h$, and is then divided by $\bar h$. Since $\bar\theta \le 1/2$, we have $\varsigma^2 \le 2\beta^2/\bar h$, and the hypothesis $8\beta^2\bar W\delta^2 \le 1$ means $4\delta^2\varsigma^2 \le \frac12$. Using $\mathrm{Y}^2 \le 2(\mathrm{Y}^{\mathrm{G}})^2 + 8\beta^4$, $|\delta| \le 1$ and the Gaussian identities $\mathbb{E}e^{a(\mathrm{Y}^{\mathrm{G}})^2} = (1-2a\varsigma^2)^{-1/2}$ and $\mathbb{E}(\mathrm{Y}^{\mathrm{G}})^4 e^{a(\mathrm{Y}^{\mathrm{G}})^2} = 3\varsigma^4(1-2a\varsigma^2)^{-5/2}$ with $a = 2\delta^2$, we find

$$\mathbb{E}_0||Y^3 \le C\bar h^{-3/2}, \qquad \mathbb{E}_0\big[\mathrm{Y}^4 e^{\delta^2\mathrm{Y}^2}\big] \le C\bar h^{-2}.$$

Using $e^{\mathrm{Y}} \le 1 + \mathrm{Y} + \frac{1}{2}\mathrm{Y}^2 e^{\mathrm{Y}}$ for $\mathrm{Y} \ge 0$, $\log(1+z) \le z$, (9.8) and (9.9), we obtain

$$\widetilde{\Psi} - \Psi_0 \le \delta^2 \mathbb{E}_0\big[\chi_q \mathrm{Y}^2\big] + C\big(|\delta|^3 \bar{h}^{-3/2} + \delta^4 \bar{h}^{-2}\big), \tag{9.10}$$

which is [77, (7.19)].

Step 5: the triangular function on the first branch. It remains to estimate $\mathbb{E}_0[\chi_q \mathrm{Y}^2]$. Let $0 = s_0 < s_1 < \cdots < s_M = q$ be the partition points in $[0, q]$. Under the Gaussian law of the increments, z^x is a Brownian motion of variance β^2 per unit time at these points, and under $\mathbb{E}_0$ its law through q has density $e^{\mathfrak{d}_x}$ with respect to it, where

$$\mathfrak{d}_x = \sum_{j=0}^{M-1} \alpha(s_j)\big(\Phi(s_{j+1}, z^x_{s_{j+1}}) - \Phi(s_j, z^x_{s_j})\big)$$

[77, (7.21)]; the masses here are the scalar masses $\alpha(s_j)$, because the block factorization of Lemma 4.20 cancels the division of the masses by the block sizes. Let $\psi_0(t) = (1 - |t-r|/\bar{h})_+$ be the triangular function supported on $[r - \bar{h}, r + \bar{h}]$, and let ∂_0 denote the derivative in the direction that shifts every coordinate z^x_t by $\psi_0(t)$. Under the Gaussian law, $\mathrm{Cov}(\mathrm{Y}_0, z^x_t) = \beta^2 \psi_0(t)$, $\mathrm{Var}(\mathrm{Y}_0) = 2\beta^2/\bar{h}$ and $\partial_0 \mathrm{Y}_0 = 2/\bar{h}$. Writing $\mathbb{E}_{\mathrm{G}}$ for the Gaussian expectation, finite-dimensional Gaussian integration by parts gives

$$\mathbb{E}_{\mathrm{G}}[\mathrm{Y}_0 f] = \beta^2 \mathbb{E}_{\mathrm{G}}[\partial_0 f], \qquad \mathbb{E}_{\mathrm{G}}[\mathrm{Y}_0^2 f] = \frac{2\beta^2}{\bar{h}} \mathbb{E}_{\mathrm{G}} f + \beta^4 \mathbb{E}_{\mathrm{G}}[\partial_0^2 f] \tag{9.11}$$

for every smooth function f of the coordinates with moderate growth. Summing by parts in $\mathfrak{d}_x$ and using $\psi_0(0) = \psi_0(q) = 0$, we have

$$\partial_0 \mathfrak{d}_x = \sum_{j=1}^{M-1} \big(\alpha(s_{j-1}) - \alpha(s_j)\big) \psi_0(s_j)\, u(s_j, z^x_{s_j}),$$
$$\partial_0^2 \mathfrak{d}_x = \sum_{j=1}^{M-1} \big(\alpha(s_{j-1}) - \alpha(s_j)\big) \psi_0(s_j)^2\, \chi(s_j, z^x_{s_j}).$$

Only the points $s_j \in (r - \bar{h}, r + \bar{h})$ contribute. As the partition contains $r - \bar{h}$ and the step function is nondecreasing, the sum of $|\alpha(s_{j-1}) - \alpha(s_j)|$ over these points is at most osc_r. Together with $|u| \le 1$ and $0 < \chi \le 1$, this implies that $|\partial_0 \mathfrak{d}_x| \le \mathrm{osc}_r$ and $|\partial_0^2 \mathfrak{d}_x| \le \mathrm{osc}_r$ [77, (7.23)].

Since $\psi_0(q) = 0$, we have $\partial_0 \chi_q = 0$. Applying (9.11) to $f = e^{\mathfrak{d}_x} \chi_q$, we obtain

$$\mathbb{E}_0[\chi_q \mathrm{Y}_0^2] = \frac{2\beta^2}{\bar{h}} \mathbb{E}_0 \chi_q + \beta^4 \mathbb{E}_0\big[\chi_q\big((\partial_0 \mathfrak{d}_x)^2 + \partial_0^2 \mathfrak{d}_x\big)\big] = \frac{2\beta^2}{\bar{h}} \mathbb{E}_0 \chi_q + \mathcal{E}_1, \qquad |\mathcal{E}_1| \le 2\beta^4 \,\mathrm{osc}_r\,. \tag{9.12}$$

With $u_r = u(r, z_r^x)$ and $\chi_r = \chi(r, z_r^x)$, we have $\partial_0 u_r = \psi_0(r)\chi_r = \chi_r$. Applying the first identity in (9.11) to $f = e^{\mathfrak{d}_x}\chi_q u_r$, we find

$$\mathbb{E}_0[\chi_q \mathrm{Y}_0 u_r] = \beta^2 \mathbb{E}_0[\chi_q \chi_r] + \beta^2 \mathbb{E}_0[\chi_q u_r\, \partial_0 \mathfrak{d}_x] = \beta^2 \mathbb{E}_0[\chi_q \chi_r] + \mathcal{E}_2, \qquad |\mathcal{E}_2| \le \beta^2 \operatorname{osc}_r . \tag{9.13}$$

These are [77, (7.24)–(7.25)]. The integrations by parts are legitimate for every fixed partition, since the scalar values grow at most linearly, their derivatives are bounded, and the density grows at most exponentially in the Gaussian variables.

Step 6: the ramp on the second branch. We condition on the whole path z^x through q. By the conditional independence in Step 1, the continuation of v after r is the scalar diffusion started from z_r^x at time r, as in Proposition 4.3(v). Its law on $[r, r+\bar h]$ has a density $e^{\mathfrak{d}_v}$ of the same form as $\mathfrak{d}_x$, over the partition points of $[r, r+\bar h]$. Let $\psi_1(t) = \min\{(t-r)_+/\bar h, 1\}$ be the ramp, and ∂_1 the corresponding derivative in the coordinates z_t^v, $t \ge r$. Under the conditional Gaussian law, $\operatorname{Cov}(\mathrm{Y}_1, z_t^v) = \beta^2 \psi_1(t)$, $\operatorname{Var}(\mathrm{Y}_1) = \beta^2/\bar h$ and $\partial_1 \mathrm{Y}_1 = 1/\bar h$. Summing by parts, $\partial_1 \mathfrak{d}_v$ consists of the terminal term $\alpha(s_{\mathrm{last}}) u(r+\bar h, z_{r+\bar h}^v)$ and internal terms whose coefficients have total absolute value at most one, so $|\partial_1 \mathfrak{d}_v| \le 2$ and $|\partial_1^2 \mathfrak{d}_v| \le 2$. By the second identity of (9.11), in conditional form, we have $\mathbb{E}_0[\mathrm{Y}_1^2 \mid z^x] = \beta^2/\bar h + \beta^4 \mathbb{E}_0[(\partial_1 \mathfrak{d}_v)^2 + \partial_1^2 \mathfrak{d}_v \mid z^x]$, and since χ_q is a function of z^x with $0 < \chi_q \le 1$,

$$\mathbb{E}_0[\chi_q \mathrm{Y}_1^2] = \frac{\beta^2}{\bar h} \mathbb{E}_0 \chi_q + \mathcal{E}_3, \qquad |\mathcal{E}_3| \le 6\beta^4. \tag{9.14}$$

For the mixed moment we need the conditional mean of Y_1. The Brownian part of $z_{r+\bar h}^v - z_r^v$ has conditional mean zero, and $u(t, Y_t^v)$ is a martingale for the diffusion started at time r, by Proposition 4.3(iv)–(v). With $\bar\alpha = \bar h^{-1} \int_r^{r+\bar h} \alpha(t)\, \mathrm{d}t$,

$$\mathbb{E}_0[\mathrm{Y}_1 \mid z^x] = \frac{\beta^2}{\bar h} \int_r^{r+\bar h} \alpha(t)\, \mathbb{E}_0\big[u(t, Y_t^v) \mid z^x\big]\, \mathrm{d}t = \beta^2 \bar\alpha\, u_r$$

[77, (7.27)]. The variables Y_0, χ_q and u_r are functions of z^x. Combining the last display with (9.13) and $|\bar\alpha - \alpha(r)| \le \operatorname{osc}_r$, we find

$$\mathbb{E}_0[\chi_q \mathrm{Y}_0 \mathrm{Y}_1] = \beta^2 \bar\alpha\, \mathbb{E}_0[\chi_q \mathrm{Y}_0 u_r] = \beta^4 \alpha(r) \mathbb{E}_0[\chi_q \chi_r] + \mathcal{E}_4, \qquad |\mathcal{E}_4| \le 2\beta^4 \operatorname{osc}_r . \tag{9.15}$$

Expanding $\mathrm{Y}^2 = (1-\bar\theta)^2 \mathrm{Y}_0^2 + 2(1-\bar\theta)\bar\theta \mathrm{Y}_0 \mathrm{Y}_1 + \bar\theta^2 \mathrm{Y}_1^2$ and using (9.12), (9.14) and (9.15), we have

$$\mathbb{E}_0[\chi_q \mathrm{Y}^2] \le \beta^2 \bar W \mathbb{E}_0 \chi_q + 2(1-\bar\theta)\bar\theta\beta^4 \alpha(r) \mathbb{E}_0[\chi_q \chi_r] + 6\beta^4(\operatorname{osc}_r + \bar\theta^2), \tag{9.16}$$

where we used $(1-\bar\theta)^2 \cdot 2\beta^2/\bar h + \bar\theta^2 \beta^2/\bar h = \beta^2 \bar W$. This is [77, (7.20)]. The triangular function contributes errors of order osc_r, while the error of order one from the ramp is multiplied by $\bar\theta^2$.

Step 7: assembly and the limit. By Proposition 4.19, applied to the perturbed path with zero multiplier and endpoint $\widetilde{Q}_N$, the constrained pressure $p_{3,n}(D)$ is at most the trial value of the perturbed path. The base path has trial value $3\mathcal{P}_\beta(\alpha)$ for $D(r,q,0)$. Combining this with (9.6), (9.7), (9.10) and the lower bound on $\mathcal{I}$, we obtain

$$p_{3,n}(D) \le 3\mathcal{P}_\beta(\alpha) + \delta^2 \mathbb{E}_0[\chi_q \mathsf{Y}^2] + C\big(|\delta|^3 \bar{h}^{-3/2} + \delta^4 \bar{h}^{-2}\big) \\ - \beta^2\delta^2\Big(\bar{W}\int_q^1 \alpha(t)\,\mathrm{d}t + \alpha(r)(2\bar{\theta} - \bar{\theta}^2) - 3\,\mathrm{osc}_r\Big) + \beta^2\delta^4\bar{W}^2.$$

We insert (9.16) and note that

$$2(1-\bar{\theta})\bar{\theta}\beta^4\alpha(r)\mathbb{E}_0[\chi_q\chi_r] - \beta^2\alpha(r)(2\bar{\theta} - \bar{\theta}^2) \\ = 2(1-\bar{\theta})\bar{\theta}\beta^2\alpha(r)\big(\beta^2\mathbb{E}_0[\chi_q\chi_r] - 1\big) - \beta^2\alpha(r)\bar{\theta}^2.$$

We discard the last term, which is nonpositive, and use $\beta^2\delta^4\bar{W}^2 \le 4\beta^2\delta^4\bar{h}^{-2}$. With $\widetilde{C} = 6\beta^4 + 3\beta^2$, this proves (9.5) for the step approximation, with $3\mathcal{P}_\beta(\alpha)$ in place of $3P_\beta$ and with $\mathbb{E}_0\chi_q - \int_q^1 \alpha(t)\,\mathrm{d}t$ in place of $\alpha(q)(q - \Gamma(q))$; this is [77, (7.29)]. The constants do not depend on the partition.

We now keep $n, r, q, \delta, \bar{h}, \bar{\theta}$ fixed and let the mesh tend to zero. The left side does not depend on the approximation. By the stability estimate Proposition 4.3(ii), the scalar values and their first two spatial derivatives converge uniformly. Driving the diffusions by the same Brownian motion, Gronwall's inequality and the uniform Lipschitz bound on the drifts show that the paths converge. The expectations $\mathbb{E}_0\chi_q$ and $\mathbb{E}_0[\chi_q\chi_r]$ converge to $\mathbb{E}\chi_q$ and $\mathbb{E}[\chi_q\chi_r]$ for the scalar diffusion Y of the minimizer, $\int_q^1 \alpha$ and $\mathcal{P}_\beta(\alpha)$ converge to their values for the minimizer, the latter being P_β, and the values at r and q are unchanged. Finally, by (4.9), we have $\mathbb{E}\chi_q - \int_q^1 \alpha(t)\,\mathrm{d}t = \alpha(q)(q - \Gamma(q))$. This proves (9.5). □

We take the limit in Step 7 at fixed $\bar{h}$ and $\bar{\theta}$; no approximation rate uniform in these parameters is needed. Only the values of α at $r \pm \bar{h}$ enter; no regularity of α at the distant overlap q is used, and q may be close to one. We now choose the parameters.

Lemma 9.5 (Separated overlaps)**.** *Let $\mathcal{J} \subset (0, q_\beta)$ be a compact interval and $d_0 > 0$. There exist $\nu_{\mathrm{sep}}, \delta_{\mathrm{sep}}, c_{\mathrm{sep}} > 0$, depending only on β, $\mathcal{J}$ and d_0, such that*

$$p_{3,n}\big(D(r,q,\delta)\big) \le 3P_\beta - c_{\mathrm{sep}}\delta^2 \tag{9.17}$$

for every $n \ge 2$ and every feasible $D(r,q,\delta)$ with $r \in \mathcal{J}$, $r + d_0 \le q \le 1$, $|q - \Gamma(q)| \le \nu_{\mathrm{sep}}$ and $|\delta| \le \delta_{\mathrm{sep}}$.

Proof. We may assume $d_0 \le 1$. Let $r_- = \min \mathcal{J}$ and $r_+ = \max \mathcal{J}$. By Corollary 4.8, α is smooth on $[0, q_\beta)$ and in particular Lipschitz, with some constant L_α, on the compact

interval $\mathcal{J}' = [r_-/2, (r_+ + q_\beta)/2] \subset (0, q_\beta)$. Let c_χ be the constant of Lemma 9.3, set $c_{\mathrm{mix}} = \beta^2 \alpha(r_-) c_\chi$, which is positive by Corollary 4.8, and let $\widetilde{C}, \widehat{C}$ be the constants of Lemma 9.4. We choose, in this order,

$$\begin{aligned}
&\bar\theta \in (0, 1/2] \text{ with } \widetilde{C}\bar\theta \le \tfrac{1}{8} c_{\mathrm{mix}} d_0;\\
&\bar h \in (0,1] \text{ with } \bar h \le \min\{\tfrac12 r_-, \tfrac12 (q_\beta - r_+), \tfrac12 d_0\} \text{ and } 2\widetilde{C} L_\alpha \bar h \le \tfrac18 c_{\mathrm{mix}} \bar\theta d_0;\\
&\nu_{\mathrm{sep}} > 0 \text{ with } 2\beta^2 \nu_{\mathrm{sep}} \le \tfrac18 c_{\mathrm{mix}} \bar\theta d_0 \bar h;\\
&\delta_{\mathrm{sep}} \in (0,1] \text{ with } 16\beta^2 \delta_{\mathrm{sep}}^2 \le \bar h \text{ and } \widehat{C}\big(\delta_{\mathrm{sep}} \bar h^{-3/2} + \delta_{\mathrm{sep}}^2 \bar h^{-2}\big) \le \tfrac18 c_{\mathrm{mix}} \bar\theta d_0.
\end{aligned}$$

Let $D = D(r, q, \delta)$ be as in the lemma. Then $\bar h < r$ and $q - r \ge d_0 > \bar h$, and $\bar W \le 2/\bar h$ because $\bar\theta \le 1/2$. We have $8\beta^2 \bar W \delta^2 \le 16\beta^2\delta^2/\bar h \le 1$, and Lemma 9.4 applies. We bound the terms in the braces of (9.5). Since $\alpha(q) \le 1$ and $|q - \Gamma(q)| \le \nu_{\mathrm{sep}}$, the first term is at most $2\beta^2\nu_{\mathrm{sep}}/\bar h$. By Lemma 9.3, $2(1-\bar\theta) \ge 1$, $\alpha(r) \ge \alpha(r_-)$ and $q - r \ge d_0$, the second term is at most $-\bar\theta\beta^2\alpha(r_-)c_\chi d_0 = -c_{\mathrm{mix}}\bar\theta d_0$. Moreover $\mathrm{osc}_r \le 2L_\alpha \bar h$, because $[r - \bar h, r + \bar h] \subset \mathcal{J}'$. Altogether,

$$p_{3,n}(D) \le 3P_\beta + \delta^2 \Big\{ \frac{2\beta^2\nu_{\mathrm{sep}}}{\bar h} - c_{\mathrm{mix}}\bar\theta d_0 + \widetilde{C}\big[\alpha(r+\bar h) - \alpha(r-\bar h) + \bar\theta^2\big] \Big\} + \widehat{C}\big(|\delta|^3 \bar h^{-3/2} + \delta^4 \bar h^{-2}\big). \quad (9.18)$$

By the choices above, the first term in the braces, the terms $2\widetilde{C}L_\alpha\bar h$ and $\widetilde{C}\bar\theta^2$, and the last term divided by δ^2 are each at most $\frac18 c_{\mathrm{mix}}\bar\theta d_0$. This proves (9.17) with $c_{\mathrm{sep}} = \frac12 c_{\mathrm{mix}}\bar\theta d_0$. □

We chose the constants in the order d_0, then $\bar\theta$, then $\bar h$, then ν_{sep}, then δ_{sep}. The condition $|\delta| \le \delta_{\mathrm{sep}}$ cannot be dropped: (9.17) is a statement about small overlap imbalances. Decreasing ν_{sep} later only shrinks the domain of (9.17).

9.3 Families of constraints

We now fix the data of the simultaneous estimate. All of the following are chosen before $n \to \infty$, in the order listed.

(F1) Compact intervals $\mathcal{J} \subset \operatorname{int} I$ with $I \subset (0, q_\beta)$, and a nonempty finite set $\mathfrak{B} \subset (0,1)$ of large-overlap cutoffs. In the dynamical argument, $\mathcal{J} = J_\beta$ and $I = I_\beta$ are the intervals of (10.13), and $\mathfrak{B} = \{q_*, b_1\}$; the fixed-scale exceptional sets of Section 10 are then subsets of the set $\mathcal{W}_n$ defined below.

(F2) The band ratio ϑ and amplitude E of Section 8.5, the constant m_β for I, a coefficient $0 < g < m_\beta E^2 G(\vartheta, E)$, and the constants d_0, S, ρ_0 of Lemma 8.7 for $\mathcal{J}$, I, ϑ, E and g.

(F3) The constants $\nu_{\mathrm{sep}}, \delta_{\mathrm{sep}}, c_{\mathrm{sep}}$ of Lemma 9.5 for $\mathcal{J}$ and d_0, and, for each $b_* \in \mathfrak{B}$, the constants $\underline{b}(b_*), \nu_{\mathrm{fix}}(b_*), c_{\mathrm{fix}}(b_*)$ of Lemma 9.2.

(F4) The tolerance and the two deficits

$$\nu = \min\Big\{\nu_{\mathrm{sep}}, \tfrac{1}{2}c_{\mathrm{tail}}, \min_{b_*\in\mathfrak{B}} \nu_{\mathrm{fix}}(b_*)\Big\}, \qquad c_{\mathrm{fix}} = \min_{b_*\in\mathfrak{B}} c_{\mathrm{fix}}(b_*), \qquad c_{\mathrm{pair}} = \tfrac{1}{2}\nu^2,$$

with c_{tail} from Lemma 9.1.

(F5) The fraction $\varkappa \in (0,1)$ of the local deficit that is retained in the disorder event.

Let (ρ_n) and (η_n) be deterministic positive sequences with $\rho_n \to 0$, and write $\varepsilon_n = E\rho_n$. We define four deterministic families of feasible overlap matrices, the local family, the separated-overlap family, the fixed-scale family, and the excluded pairs:

$$\begin{aligned}
\mathcal{D}_n^{\mathrm{loc}} &= \big\{D(r,q,\delta) : r\in\mathcal{J},\ \rho_n \le q-r \le d_0,\ \varepsilon_n \le |\delta| \le 2\varepsilon_n\big\},\\
\mathcal{D}_n^{\mathrm{sep}} &= \big\{D(r,q,\delta) : r\in\mathcal{J},\ q-r > d_0,\ |q-\Gamma(q)| \le \nu,\ \varepsilon_n \le |\delta| \le 2\varepsilon_n\big\},\\
\mathcal{D}_n^{\mathrm{fix}} &= \bigcup_{b_*\in\mathfrak{B}} \big\{D(r,q,\delta) : q\ge b_*,\ |r\pm\delta| \le \underline{b}(b_*),\ |q-\Gamma(q)| \le \nu,\ |\delta| \ge \eta_n\big\},\\
\mathcal{D}_n^{\mathrm{pair}} &= \big\{D_2(\bar q) : \big||\bar q| - \Gamma(|\bar q|)\big| > \nu\big\},
\end{aligned}$$

where only feasible matrices are included. There are at most $(n+1)^3$ matrices in $\mathcal{D}_n^{\mathrm{loc}} \cup \mathcal{D}_n^{\mathrm{sep}}$, at most $|\mathfrak{B}|(n+1)^3$ in $\mathcal{D}_n^{\mathrm{fix}}$, and at most $n+1$ in $\mathcal{D}_n^{\mathrm{pair}}$. Since $\nu < c_{\mathrm{tail}}$, the family $\mathcal{D}_n^{\mathrm{pair}}$ contains every feasible $D_2(\bar q)$ with $|\bar q| \ge q_{\mathrm{tail}}$. For a family $\mathcal{D}$ of $\ell\times\ell$ matrices, we write $\{R\in\mathcal{D}\}$ for the set of ℓ-tuples whose overlap matrix lies in $\mathcal{D}$. The corresponding sets of configurations are the narrow triples, the fixed-scale triples and the excluded pairs,

$$\begin{aligned}
\mathcal{N}_n &= \big\{(v,x,y)\in\Sigma_n^3 : r\in\mathcal{J},\ q-r\ge\rho_n,\ \varepsilon_n \le |\delta| \le 2\varepsilon_n\big\},\\
\mathcal{W}_n &= \bigcup_{b_*\in\mathfrak{B}} \big\{(v,x,y)\in\Sigma_n^3 : R(x,y)\ge b_*,\ |R(v,x)|, |R(v,y)| \le \underline{b}(b_*),\\
&\qquad\qquad |R(v,x)-R(v,y)| \ge 2\eta_n\big\},\\
\mathcal{P}_n &= \{R\in\mathcal{D}_n^{\mathrm{pair}}\} = \big\{(x,y)\in\Sigma_n^2 : \big||R(x,y)| - \Gamma(|R(x,y)|)\big| > \nu\big\},
\end{aligned}$$

with r,δ,q the coordinates of the triple.

Lemma 9.6 (Covering)**.** *If $(v,x,y)\in\mathcal{N}_n$, then $(x,y)\in\mathcal{P}_n$ or the overlap matrix of (v,x,y) lies in $\mathcal{D}_n^{\mathrm{loc}}\cup\mathcal{D}_n^{\mathrm{sep}}$. If $(v,x,y)\in\mathcal{W}_n$, then $(x,y)\in\mathcal{P}_n$ or the overlap matrix of (v,x,y) lies in $\mathcal{D}_n^{\mathrm{fix}}$.*

Proof. Let $(v,x,y)\in\mathcal{N}_n$. Then $q = R(x,y) \ge r+\rho_n > 0$, since $r\in\mathcal{J}\subset(0,1)$. If $|q-\Gamma(q)| > \nu$, then $(x,y)\in\mathcal{P}_n$. Otherwise $D(r,q,\delta)$ lies in $\mathcal{D}_n^{\mathrm{loc}}$ if $q-r\le d_0$ and in $\mathcal{D}_n^{\mathrm{sep}}$ if $q-r>d_0$. Next let $(v,x,y)\in\mathcal{W}_n$, for the cutoff $b_*\in\mathfrak{B}$. Then $q = R(x,y)\ge b_* > 0$, $|r\pm\delta| \le \underline{b}(b_*)$ and $|\delta|\ge\eta_n$, so either $(x,y)\in\mathcal{P}_n$ or $D(r,q,\delta)\in\mathcal{D}_n^{\mathrm{fix}}$. □

By Lemma 9.6, every narrow triple is covered by the local regime, the regime of separated overlaps, or the pair regime. The local regime uses the cubic deficit of Section 7

through Lemma 8.7; the other two use the quadratic deficits of Lemmas 9.1 and 9.5. No fractional moment bound at a distant overlap q is needed, and the construction of Section 7 is used only for $q - r \le d_0$.

9.4 The simultaneous event

The conversion of constrained-pressure bounds to Gibbs probabilities is treated in [77, Lemma 4.2], where the logarithm of the Gibbs probability is concentrated about its mean. Here we bound the constrained numerators and the full partition-function denominator separately. For the local numerators we use the positive fractional-moment mechanism of [8, Section 2.3], through Lemma 8.7. This controls their deviations at the required shrinking scale. The simultaneous event below combines this with a variance bound for the denominator and retains separate disorder and conditional Gibbs errors at the shrinking scales.

We impose the following conditions on the scales. Recall that $\Delta_{n,\beta} = P_\beta - p_n(\beta)$.

$$\limsup_{n\to\infty} \frac{\Delta_{n,\beta}}{\rho_n^3} < \frac{\varkappa g}{3}, \tag{S1}$$

$$\lim_{n\to\infty} \frac{n^{7/15}}{n^2\rho_n^6} = 0, \tag{S2}$$

$$\lim_{n\to\infty} \frac{n\rho_n^4}{\log n} = \infty, \tag{S3}$$

$$\lim_{n\to\infty} \frac{n\eta_n^4}{\log n} = \infty, \tag{S4}$$

$$\lim_{n\to\infty} \frac{\rho_n^3}{\eta_n^2} = 0. \tag{S5}$$

Condition (S1) says that the mean bias is absorbed by the retained part of the cubic deficit, with a strict constant margin. Given the variance bound $\operatorname{Var} F_n \le C_v n^{7/15}$ of Theorem 4.10, (S2) makes the Chebyshev bound for the denominator tend to zero. Conditions (S3) and (S4) make the numerator bounds summable over polynomially many constraints, and (S5) makes the loss in the denominator negligible against the fixed-scale deficit. By (S1), we may fix a_0 and c_0 with

$$\limsup_{n\to\infty} \frac{\Delta_{n,\beta}}{\rho_n^3} < a_0 < c_0 < \frac{\varkappa g}{3}, \tag{9.19}$$

for instance $c_0 = \frac{1}{2}(a_0 + \varkappa g/3)$.

We define $\mathcal{A}_n$ as the set of disorders J for which all of the following hold:

(A1) $F_n \ge n(P_\beta - c_0\rho_n^3)$;

(A2) $n^{-1}\log Z_{3,n}(D) \le 3P_\beta - \varkappa g\rho_n^3$ for every $D \in \mathcal{D}_n^{\mathrm{loc}} \cup \mathcal{D}_n^{\mathrm{sep}}$;
(A3) $n^{-1}\log Z_{3,n}(D) \le 3P_\beta - \frac{1}{2}c_{\mathrm{fix}}\eta_n^2$ for every $D \in \mathcal{D}_n^{\mathrm{fix}}$;
(A4) $n^{-1}\log Z_{2,n}(D) \le 2P_\beta - \frac{1}{2}c_{\mathrm{pair}}$ for every $D \in \mathcal{D}_n^{\mathrm{pair}}$.

The event $\mathcal{A}_n$ is measurable, since all constrained partition functions are finite sums of continuous functions of the disorder.

Lemma 9.7 (Simultaneous static event)**.** *Fix the data* (F1)–(F5)*, and let* (ρ_n)*,* (η_n) *satisfy* (S1)–(S5)*. There exist* $C, c > 0$ *and* n_0*, depending only on* β*, on the data* (F1)–(F5) *and on the sequences, such that for all* $n \ge n_0$*,*

$$\mathbb{P}(\mathcal{A}_n^c) \le C\frac{n^{7/15}}{n^2\rho_n^6} + C(n+1)^3\big(e^{-cn\rho_n^4} + e^{-cn\eta_n^4}\big) + C(n+1)e^{-cn}, \tag{9.20}$$

and the right side tends to zero. Moreover, for all $n \ge n_0$ *and every disorder* $J \in \mathcal{A}_n$*,*

$$\begin{gathered}(\pi^J)^{\otimes 3}\big(\{R \in \mathcal{D}_n^{\mathrm{loc}} \cup \mathcal{D}_n^{\mathrm{sep}}\}\big) \le C(n+1)^3 e^{-cn\rho_n^3}, \qquad (\pi^J)^{\otimes 3}(\mathcal{N}_n) \le C(n+1)^3 e^{-cn\rho_n^3},\\ (\pi^J)^{\otimes 3}\big(\{R \in \mathcal{D}_n^{\mathrm{fix}}\}\big) \le C(n+1)^3 e^{-cn\eta_n^2}, \qquad (\pi^J)^{\otimes 3}(\mathcal{W}_n) \le C(n+1)^3 e^{-cn\eta_n^2},\\ (\pi^J)^{\otimes 2}(\mathcal{P}_n) \le C(n+1)e^{-cn}.\end{gathered} \tag{9.21}$$

The same bounds hold for the images of these sets under every element of $\mathfrak{O}_3$*, respectively* $\mathfrak{O}_2$*, and the set of triples* (v,x,y) *with* $(x,y) \in \mathcal{P}_n$ *has* $(\pi^J)^{\otimes 3}$*-probability* $(\pi^J)^{\otimes 2}(\mathcal{P}_n)$*.*

Proof. We bound the probability that each of (A1)–(A4) fails, take a union bound over the deterministic families, and then divide the bounds on the numerators by the bound on the denominator, on $\mathcal{A}_n$. Let n_0 be so large that, for $n \ge n_0$, we have $\Delta_{n,\beta} \le a_0\rho_n^3$, $\operatorname{Var} F_n \le C_v n^{7/15}$, $\rho_n \le \min\{\rho_0, 1\}$, $2\varepsilon_n \le \delta_{\mathrm{sep}}$, $\varkappa g\rho_n \le \frac{1}{2}c_{\mathrm{sep}}E^2$, $12c_0\rho_n^3 \le c_{\mathrm{fix}}\eta_n^2$ and $8c_0\rho_n^3 \le c_{\mathrm{pair}}$. This is possible by (9.19), Theorem 4.10, $\rho_n \to 0$ and (S5).

Step 1: the denominator. Since $\mathbb{E}F_n = n(P_\beta - \Delta_{n,\beta}) \ge n(P_\beta - a_0\rho_n^3)$, the failure of (A1) implies $F_n - \mathbb{E}F_n < -n(c_0 - a_0)\rho_n^3$. By Chebyshev's inequality,

$$\mathbb{P}\big(F_n < n(P_\beta - c_0\rho_n^3)\big) \le \frac{C_v n^{7/15}}{(c_0 - a_0)^2 n^2 \rho_n^6}.$$

Step 2: local numerators. Let $D \in \mathcal{D}_n^{\mathrm{loc}}$ and $s = S\rho_n$. By Markov's inequality applied to $Z_{3,n}(D)^s$ and by (8.19),

$$\mathbb{P}\big(n^{-1}\log Z_{3,n}(D) > 3P_\beta - \varkappa g\rho_n^3\big) \le \mathbb{E}Z_{3,n}(D)^s\, e^{-ns(3P_\beta - \varkappa g\rho_n^3)} \le e^{-(1-\varkappa)gS\, n\rho_n^4}.$$

This is the only place where a fractional moment enters. A Gaussian concentration bound at a mean deficit of order ρ_n^3 would give the exponent $n\rho_n^6$, which is $R^6 n^{-2/5}$ at the scale $\rho_n = Rn^{-7/30}$ and tends to zero.

Step 3: the other numerators. By Remark 4.22, $\log Z_{\ell,n}(D)$ is Lipschitz in $(J_{ij})_{i<j}$ with constant $\beta\ell\sqrt{(n-1)/2}$. By the Gaussian concentration inequality (4.19) of Proposition 4.21, for every $t > 0$,

$$\mathbb{P}\big(\log Z_{\ell,n}(D) - \mathbb{E}\log Z_{\ell,n}(D) \geq nt\big) \leq 2\exp\Big(-\frac{n^2t^2}{\beta^2\ell^2(n-1)}\Big) \leq 2\exp\Big(-\frac{nt^2}{\beta^2\ell^2}\Big). \quad (9.22)$$

For $D \in \mathcal{D}_n^{\mathrm{sep}}$, we have $\varepsilon_n \leq |\delta| \leq 2\varepsilon_n \leq \delta_{\mathrm{sep}}$, and Lemma 9.5 implies that $p_{3,n}(D) \leq 3P_\beta - c_{\mathrm{sep}}E^2\rho_n^2$. Since $\varkappa g\rho_n^3 \leq \frac{1}{2}c_{\mathrm{sep}}E^2\rho_n^2$, the threshold in (A2) exceeds the mean $p_{3,n}(D)$ by at least $\frac{1}{2}c_{\mathrm{sep}}E^2\rho_n^2$. For $D \in \mathcal{D}_n^{\mathrm{fix}}$, we see from Lemma 9.2 and $|\delta| \geq \eta_n$ that $p_{3,n}(D) \leq 3P_\beta - c_{\mathrm{fix}}\eta_n^2$, and the threshold in (A3) exceeds the mean by at least $\frac{1}{2}c_{\mathrm{fix}}\eta_n^2$. For $D \in \mathcal{D}_n^{\mathrm{pair}}$, we have $p_{2,n}(D) < 2P_\beta - \frac{1}{2}\nu^2 = 2P_\beta - c_{\mathrm{pair}}$ by Lemma 9.1, and the threshold in (A4) exceeds the mean by at least $\frac{1}{2}c_{\mathrm{pair}}$. Applying (9.22) with $\ell = 3, 3, 2$ and these values of t, we obtain, for a single matrix of the respective family, the failure probabilities

$$2e^{-c_{\mathrm{sep}}^2E^4n\rho_n^4/(36\beta^2)}, \qquad 2e^{-c_{\mathrm{fix}}^2n\eta_n^4/(36\beta^2)}, \qquad 2e^{-c_{\mathrm{pair}}^2n/(16\beta^2)}.$$

Step 4: the union bound. Recall that the families are deterministic and that their sizes were bounded in Section 9.3. Summing the bounds of Steps 1–3 over them, we have

$$\mathbb{P}(\mathcal{A}_n^c) \leq \frac{C_v}{(c_0-a_0)^2}\frac{n^{7/15}}{n^2\rho_n^6} + (n+1)^3\Big(e^{-(1-\varkappa)gSn\rho_n^4} + 2e^{-c_{\mathrm{sep}}^2E^4n\rho_n^4/(36\beta^2)}\Big) + 2|\mathfrak{B}|(n+1)^3e^{-c_{\mathrm{fix}}^2n\eta_n^4/(36\beta^2)} + 2(n+1)e^{-c_{\mathrm{pair}}^2n/(16\beta^2)},$$

which is (9.20). The first term tends to zero by (S2), and the others by (S3) and (S4), since $(n+1)^3e^{-cn\rho_n^4} = \exp\{3\log(n+1) - cn\rho_n^4\}$, and similarly for η_n.

Step 5: Gibbs probabilities on $\mathcal{A}_n$. Fix $J \in \mathcal{A}_n$. By (A1), $Z_n^\ell \geq \exp\{\ell n(P_\beta - c_0\rho_n^3)\}$. The ratio $Z_{\ell,n}(D)/Z_n^\ell$ is the Gibbs probability of $\{R = D\}$. For $D \in \mathcal{D}_n^{\mathrm{loc}} \cup \mathcal{D}_n^{\mathrm{sep}}$, (A2) gives

$$\frac{Z_{3,n}(D)}{Z_n^3} \leq e^{-n(\varkappa g - 3c_0)\rho_n^3},$$

and the coefficient $\varkappa g - 3c_0$ is positive by (9.19). For $D \in \mathcal{D}_n^{\mathrm{fix}}$ and $D \in \mathcal{D}_n^{\mathrm{pair}}$, (A3), (A4) and the choice of n_0 give, respectively,

$$\frac{Z_{3,n}(D)}{Z_n^3} \leq e^{-n(c_{\mathrm{fix}}\eta_n^2/2 - 3c_0\rho_n^3)} \leq e^{-c_{\mathrm{fix}}n\eta_n^2/4}, \qquad \frac{Z_{2,n}(D)}{Z_n^2} \leq e^{-n(c_{\mathrm{pair}}/2 - 2c_0\rho_n^3)} \leq e^{-c_{\mathrm{pair}}n/4}.$$

Summing over the families gives the first three bounds of (9.21), with $c = \min\{\varkappa g - 3c_0, \frac{1}{4}c_{\mathrm{fix}}, \frac{1}{4}c_{\mathrm{pair}}\}$ and $C = |\mathfrak{B}|$. An event that depends only on (x, y) has the same probability under $(\pi^J)^{\otimes 3}$ as under $(\pi^J)^{\otimes 2}$, which is the last assertion of the lemma. By

Lemma 9.6, $\mathcal{N}_n$ is contained in the union of $\{R \in \mathcal{D}_n^{\mathrm{loc}} \cup \mathcal{D}_n^{\mathrm{sep}}\}$ and the set of triples with $(x,y) \in \mathcal{P}_n$, so

$$(\pi^J)^{\otimes 3}(\mathcal{N}_n) \le (n+1)^3 e^{-n(\varkappa g - 3c_0)\rho_n^3} + (n+1)e^{-c_{\mathrm{pair}}n/4},$$

and similarly $(\pi^J)^{\otimes 3}(\mathcal{W}_n) \le |\mathfrak{B}|(n+1)^3 e^{-c_{\mathrm{fix}}n\eta_n^2/4} + (n+1)e^{-c_{\mathrm{pair}}n/4}$. We may assume $\eta_n \le 1$, since otherwise $\mathcal{D}_n^{\mathrm{fix}}$ and $\mathcal{W}_n$ are empty. As also $\rho_n \le 1$, the term $e^{-c_{\mathrm{pair}}n/4}$ is at most $e^{-c_{\mathrm{pair}}n\rho_n^3/4}$ and at most $e^{-c_{\mathrm{pair}}n\eta_n^2/4}$. We obtain the last two bounds of (9.21) with $C = |\mathfrak{B}| + 1$. The statements for the images under $\mathfrak{O}_3$ and $\mathfrak{O}_2$ follow from (9.1). □

Estimate	Bound	Kind	Theorem 1.1
(A1), denominator	$n^{7/15}/(n^2\rho_n^6)$	disorder	$n^{-2/15}$
(A2), local numerators	$(n+1)^3 e^{-cn\rho_n^4}$	disorder	$e^{-cn^{1/15}}$
(A2), separated overlaps	$(n+1)^3 e^{-cn\rho_n^4}$	disorder	$e^{-cn^{1/15}}$
(A3), fixed-scale numerators	$(n+1)^3 e^{-cn\eta_n^4}$	disorder	$e^{-cn^{1/21}}$
(A4), pair numerators	$(n+1)e^{-cn}$	disorder	e^{-cn}
narrow triples	$(n+1)^3 e^{-cn\rho_n^3}$	Gibbs, on $\mathcal{A}_n$	$e^{-cn^{3/10}}$
fixed-scale triples	$(n+1)^3 e^{-cn\eta_n^2}$	Gibbs, on $\mathcal{A}_n$	$e^{-cn^{11/21}}$
excluded pairs	$(n+1)e^{-cn}$	Gibbs, on $\mathcal{A}_n$	e^{-cn}

Table 9.1. The error terms of Lemma 9.7, up to constant factors, at the scales $\rho_n = Rn^{-7/30}$ and $\eta_n = n^{-5/21}$ of Theorem 1.1. The first five rows bound the disorder probability $\mathbb{P}(\mathcal{A}_n^c)$; they are paid once. The last three rows are Gibbs probabilities at a fixed disorder in $\mathcal{A}_n$; only these are multiplied by counts in Sections 10 to 12. The last column omits polynomial factors.

Remark 9.8 (The scales of the barrier theorem). Fix the data (F1)–(F5), and write C_b for the constant of the upper bound on the mean bias that is used.

(a) By Theorem 4.11, $0 \le \Delta_{n,\beta} \le C_b n^{-7/10}$ for $n \ge 2$. We choose

$$\rho_n = Rn^{-7/30}, \qquad \varepsilon_n = E\rho_n, \qquad \eta_n = n^{-5/21}, \qquad s_n = S\rho_n, \qquad R^3 > \frac{3C_b}{\varkappa g}. \tag{9.23}$$

Then $\Delta_{n,\beta}/\rho_n^3 \le C_b/R^3 < \varkappa g/3$, which is (S1). Moreover $n^{7/15}/(n^2\rho_n^6) = R^{-6}n^{-2/15}$, $n\rho_n^4 = R^4 n^{1/15}$, $n\eta_n^4 = n^{1/21}$ and $\rho_n^3/\eta_n^2 = R^3 n^{-47/210}$, which proves (S2)–(S5). The Gibbs probabilities decay on the scales $n\rho_n^3 = R^3 n^{3/10}$ and $n\eta_n^2 = n^{11/21}$.

(b) More generally, let $\rho_n = Rn^{-\lambda}$ and $\eta_n = n^{-\xi}$, where the scale exponents λ and ξ satisfy (10.10). Suppose the mean-bias bound (10.11) and $R^3 > 3C_b/(\varkappa g)$. Then $\Delta_{n,\beta}/\rho_n^3 \le C_b/R^3 < \varkappa g/3$ for large n, which is (S1). We have (S2) because $n^{7/15}/(n^2\rho_n^6) = R^{-6}n^{7/15-2+6\lambda}$ and $6\lambda < 3/2 < 23/15$. We have (S3) and (S4) because $1 - 4\lambda > 0$ and $1 - 4\xi > 0$, and (S5) holds because $\rho_n^3/\eta_n^2 = R^3 n^{2\xi - 3\lambda}$. Part (a) is the case $(\lambda, \xi) = (7/30, 5/21)$, with Theorem 4.11 as the bound (10.11).

The dynamical argument needs in addition the conditions (11.1): $\eta_n/\rho_n \to E'$ for a fixed $E' \in [0, E/10]$, and $n^{-1} = o(\varepsilon_n)$. In part (b) we have $E' = 0$, since $\eta_n/\varepsilon_n = (ER)^{-1} n^{\lambda-\xi}$ and $\lambda < \xi$; in part (a), this ratio is $(ER)^{-1} n^{-1/210}$. Also $1/(n\varepsilon_n) = (ER)^{-1} n^{\lambda-1} \to 0$. The conditions (S1)–(S3) do not involve η_n. For $\eta_n = \rho_n/1000$, as in Section 12.7, where $E' = 1/1000 = E/10$ because $E = 1/100$, conditions (S4) and (S5) follow from (S3) and $\rho_n \to 0$, since $n\eta_n^4 = 10^{-12} n\rho_n^4$ and $\rho_n^3/\eta_n^2 = 10^6 \rho_n$. The variance input $n^{7/15}$ of Theorem 4.10 and the profile cost s^5 of Lemma 8.5 do not depend on the choice of the scales, and $s_n^5/\rho_n^3 = S^5 R^2 n^{-2\lambda}$. Within this argument the exponent λ is limited both by one third of the exponent in the upper bound on the mean bias and by $1/4$, the latter through (S3).

Remark 9.9 (How the estimate is used). The later sections use Lemma 9.7 in the following form. If a bad event at a fixed disorder is a union of at most K_n events, each contained in one of the sets of (9.21), in one of their images under $\mathfrak{O}_3$, or in a lift of $\mathcal{P}_n$ or of one of its images to more configurations, then its probability under the joint law of the disorder and the Gibbs samples is at most

$$\mathbb{P}(\mathcal{A}_n^c) + K_n \sup_{J \in \mathcal{A}_n} \left\{ (\pi^J)^{\otimes 3}(\mathcal{N}_n) + (\pi^J)^{\otimes 3}(\mathcal{W}_n) + (\pi^J)^{\otimes 2}(\mathcal{P}_n) \right\}$$

up to a fixed factor for the orientations. The count K_n multiplies only the Gibbs probabilities. For example, with $K_n = e^{CL_n}$ and $L_n = An^{7/30}$, we have $K_n (n+1)^3 e^{-cR^3 n^{3/10}} \to 0$ for every fixed A and C. No such count may be applied to the disorder probability $\mathbb{P}(\mathcal{A}_n^c)$, for which this section gives only the bound $O(n^{-2/15})$ at the scales of Remark 9.8(a). In Section 10, $\mathcal{A}_n$ is intersected with the disorder event of Lemma 10.1; the intersection has probability tending to one by a union bound, and no independence between the two events is needed.

PART III

Dynamics: proof of the lower bound

SECTION 10

Overlap neighborhoods and reference configurations

This section collects the probabilistic input of the dynamical argument in Sections 11 and 12. It has three parts. In Section 10.1, we prove Lemma 10.1, the overlap-neighborhood mass bound. It states that outside an event of vanishing disorder probability, for every set $S \subseteq \Sigma_n$, every $x \in S$ outside a set of Gibbs mass less than $e^{-D_1 L_n}$ has the property that the configurations $z \in S$ with $|R(x,z)| \geq q_*$ carry Gibbs mass at least $e^{-D_3 L_n}$; here $D_1 > 0$ is arbitrary, and D_3 depends only on β and D_1. This is the argument of [77, Proposition 6.1], carried out at every diverging sublinear scale L_n. In Section 10.2, we fix the parameters of the dynamical argument and the order in which they are chosen (Table 10.1). We then define the set $\mathcal{B}_n$ of exceptional triples as an explicit union and deduce from Lemma 9.7 that $\pi^{\otimes 3}(\mathcal{B}_n)$ is small on one disorder event (Lemma 10.2). In Section 10.3, we combine these two inputs. At each disorder in an event of probability tending to one, we fix one family of Gibbs reference configurations and one set V of small Gibbs mass such that the four properties used in Section 11 hold along every path that avoids V (Lemma 10.4).

The section uses Lemma 9.7 and the pressure estimate (9.3) from Section 9, the scalar constant of (5.20) from Section 5, the band of (7.37) from Section 7, and the imported results Propositions 4.21, 4.27, and 4.28. Lemma 10.4 is the only point at which the static estimates of Part II enter the induction of Section 11; its bound on $\pi(V)$ enters the stationary sign test of Section 12. Although Theorem 1.1 uses only the scale exponent $7/30$, we state the construction for general exponents, because this shows which conditions on the scales each step uses (see Table 12.2).

Throughout this section, $\beta > 1$ is fixed, and we write $\pi = \pi^J$. For a vector $\bar{J} = (\bar{J}_{ij})_{1 \leq i < j \leq n}$ of real numbers, we write $H_n^{\bar{J}}$ and $Z_n^{\bar{J}}$ for the Hamiltonian and the partition function in which J is replaced by $\bar{J}$. For $k \geq 1$ and a set $\mathcal{U} \subseteq \Sigma_n^k$, $\pi^{\otimes k}(\mathcal{U})$ is the probability, at fixed disorder, that k independent samples from π form a k-tuple in $\mathcal{U}$; we also write $\pi^{\otimes 2}(|R| \geq q)$ for $\pi^{\otimes 2}(\{(x,y) : |R(x,y)| \geq q\})$. The Gibbs measure is invariant under the spin flip $x \mapsto -x$, because $H_n^J(-x) = H_n^J(x)$, and $\pi^{\otimes k}$ is invariant under independently reversing whole replicas and under permutations of the replicas. All constants depend on β and on the fixed parameters listed in Table 10.1, and never on n or on the disorder.

10.1 Overlap neighborhoods at sublinear scales

The bound below was proved in [77, Proposition 6.1] at the scale $L_n = n^{1/10000}$ used there; this is Proposition 4.29. Its proof uses the scale only through the conditions $L_n \to \infty$ and $L_n = o(n)$. Our time scale is different, and the constants of the bound enter the construction of Section 10.3. For these reasons we state the extension and write out the proof. The ingredients are those of [77]: the upper tail of the energy of a Gibbs sample (Proposition 4.27, whose proof in [77, Lemma 5.1] uses the Brascamp–Lieb inequality [30] in the form of [77, Lemma 2.1]), the disorder rotation estimate (Proposition 4.28; [77, Lemma 5.2]), Gaussian concentration for Lipschitz functions (Proposition 4.21; [77, Lemma 2.1]), and the Paley–Zygmund inequality.

Lemma 10.1 (Overlap-neighborhood mass bound at sublinear scales)**.** *There exists $q_* = q_*(\beta) \in (0,1)$ with the following property. Let (L_n) be a deterministic sequence of positive numbers with $L_n \to \infty$ and $L_n/n \to 0$, and let $D_1 > 0$. Then there exist $D_3 > 0$ and $D_4 > 0$, depending only on β and D_1, and disorder events $\mathcal{C}_n(D_1)$ such that $\mathbb{P}(\mathcal{C}_n(D_1)^c) \le 3e^{-D_4 L_n}$ for all sufficiently large n, and such that on $\mathcal{C}_n(D_1)$,*

$$\pi\big(\{x \in S : \pi(\{z \in S : |R(x,z)| \ge q_*\}) < e^{-D_3 L_n}\}\big) < e^{-D_1 L_n} \qquad \text{for every } S \subseteq \Sigma_n. \tag{10.1}$$

Both masses in (10.1) are masses under π itself, not under π conditioned on S. The inner mass is the probability that a single Gibbs sample lands in the overlap neighborhood $\{z \in S : |R(x,z)| \ge q_*\}$ of x in S; in Section 10.3 it is the success probability of one reference sample. The overlap q_* is chosen before D_1 and before the sequence (L_n), and D_3 and D_4 do not depend on the sequence either. Since $L_n \to \infty$, the lemma implies that $\mathbb{P}(\mathcal{C}_n(D_1)) \to 1$. The event $\mathcal{C}_n(D_1)$ and the threshold beyond which the estimates below hold depend on (L_n); the lemma does not provide one event valid for all sequences at once. For $L_n = n^{1/10000}$, Lemma 10.1 contains Proposition 4.29.

Proof. The proof has three steps. In the first, we show that outside an event of vanishing probability, every set of Gibbs mass at least $e^{-D_1 L_n}$ contains a fraction at least $e^{-D_2 L_n}$ of pairs with overlap at least q_* in absolute value. The idea, from [77, Section 6], is that if a set has few such pairs, then a second moment computation shows that a small rotation of the disorder increases the partition function restricted to this set by nearly its full Gaussian factor. In the second step, we show that this gain outweighs the loss caused by shrinking the original disorder, which the upper energy tail controls, and that the rotation estimate makes the resulting increase of $\log Z_n^J$ unlikely. In the third step, we apply the pair bound to the set of exceptional targets itself.

Constants. By Proposition 4.27, there exist $e_1 \in (0, \beta/2)$ and $c_E > 0$ such that the event

$$\mathcal{E}_n = \big\{\pi(\{x : H_n^J(x) > ne_1\}) \le e^{-c_E n}\big\}$$

satisfies $\mathbb{P}(\mathcal{E}_n^c) \le e^{-c_E n}$ for all sufficiently large n. We define

$$g_0 = \frac{\beta^2}{4} - \frac{\beta e_1}{2},$$

which is positive because $e_1 < \beta/2$. We fix $q_* \in (0,1)$ with $\beta^2 q_* < g_0/8$. These choices depend only on β. Given $D_1 > 0$, we fix constants Λ, D_2, and D_3 such that

$$\Lambda > \frac{8(D_1+1)}{g_0}, \qquad D_2 > \frac{\beta^2\Lambda}{2} + 1, \qquad D_3 > D_1 + D_2. \tag{10.2}$$

They depend only on β and D_1. Now let (L_n) be as in the statement, write $L = L_n$, and let

$$\omega = \omega_n = \frac{\Lambda L_n}{n}.$$

Then $\omega_n \to 0$ and $n\omega_n = \Lambda L_n \to \infty$. We have $0 < \omega < 1/2$ for all sufficiently large n; we assume this from now on.

Step 1: pairs with large overlap in every large set. Let $\mathcal{V}_n$ be the event that some $S \subseteq \Sigma_n$ satisfies

$$\pi(S) \ge e^{-D_1 L} \qquad \text{and} \qquad \pi(\cdot \mid S)^{\otimes 2}(|R| \ge q_*) < e^{-D_2 L}. \tag{10.3}$$

This event is measurable, since it is a finite union of events, each defined by finitely many continuous functions of the disorder. Off $\mathcal{V}_n$, we have

$$\pi(\cdot \mid S)^{\otimes 2}(|R| \ge q_*) \ge e^{-D_2 L} \qquad \text{for every } S \subseteq \Sigma_n \text{ with } \pi(S) \ge e^{-D_1 L}. \tag{10.4}$$

We will show that $\mathbb{P}(\mathcal{V}_n) \to 0$.

Fix $J \in \mathcal{E}_n \cap \mathcal{V}_n$ and a set S satisfying (10.3). We first remove the configurations of high energy from S. Let

$$S' = S \cap \{x : H_n^J(x) \le n e_1\}.$$

For large n, we have $e^{-c_E n} \le e^{-D_1 L}/2 \le \pi(S)/2$, because $D_1 L = o(n)$. This implies that

$$\pi(S') \ge \pi(S) - \pi(\{x : H_n^J(x) > n e_1\}) \ge \pi(S)/2.$$

Restricting the pairs to $S' \times S'$ and dividing by $\pi(S')^2$ instead of $\pi(S)^2$, we find

$$\pi(\cdot \mid S')^{\otimes 2}(|R| \ge q_*) \le \Big(\frac{\pi(S)}{\pi(S')}\Big)^2 \pi(\cdot \mid S)^{\otimes 2}(|R| \ge q_*) < 4e^{-D_2 L}. \tag{10.5}$$

Let $\widetilde{J} = (\widetilde{J}_{ij})_{1 \le i < j \le n}$ be a standard Gaussian vector independent of J, and define

$$W = \sum_{x \in S'} \pi(x \mid S')\, e^{\beta\sqrt{\omega} H_n^{\widetilde{J}}(x)}, \qquad J' = \sqrt{1-\omega}\, J + \sqrt{\omega}\, \widetilde{J}.$$

We write $\mathbb{P}_{\widetilde{J}}$ and $\mathbb{E}_{\widetilde{J}}$ for probability and expectation over $\widetilde{J}$ with J fixed. Recall from (3.3) that $\mathbb{E}H_n^{\widetilde{J}}(x)H_n^{\widetilde{J}}(y) = (nR(x,y)^2-1)/2$. Using $\mathbb{E}e^X = e^{\mathbb{E}X^2/2}$ for a centered Gaussian variable X, we find

$$\mathbb{E}_{\widetilde{J}}W = e^{\beta^2\omega(n-1)/4}, \qquad \frac{\mathbb{E}_{\widetilde{J}}W^2}{(\mathbb{E}_{\widetilde{J}}W)^2} = \sum_{x,y\in S'} \pi(x \mid S')\,\pi(y \mid S')\exp\Big\{\frac{\beta^2\omega}{2}\big(nR(x,y)^2-1\big)\Big\}.$$

We split the last sum according to whether $|R(x,y)| < q_*$. On the first part, the exponent is at most

$$\Theta = \frac{\beta^2 n\omega q_*^2}{2}.$$

On the second part, the exponent is at most $\beta^2 n\omega/2 = \beta^2\Lambda L/2$, and we use (10.5). Since $D_2 > \beta^2\Lambda/2 + 1$, this gives

$$\frac{\mathbb{E}_{\widetilde{J}}W^2}{(\mathbb{E}_{\widetilde{J}}W)^2} \le e^{\Theta} + 4e^{(\beta^2\Lambda/2 - D_2)L} \le e^{\Theta} + 4e^{-L} \le 2e^{\Theta}$$

once $L \ge \log 4$. The Paley–Zygmund inequality $\mathbb{P}(W \ge \mathbb{E}W/2) \ge (\mathbb{E}W)^2/(4\mathbb{E}W^2)$, valid for every nonnegative random variable W with finite second moment, then implies that

$$\mathbb{P}_{\widetilde{J}}\big(W \ge \mathbb{E}_{\widetilde{J}}W/2\big) \ge \frac{e^{-\Theta}}{8}. \tag{10.6}$$

Next, we raise this probability to $1/2$ at the cost of a lower threshold. Let $f(\widetilde{J}) = \log W$. Differentiating the finite sum defining W, we have

$$\frac{\partial f}{\partial \widetilde{J}_{ij}} = \frac{\beta}{W}\sqrt{\frac{\omega}{n}} \sum_{x\in S'} \pi(x \mid S')\, e^{\beta\sqrt{\omega}H_n^{\widetilde{J}}(x)} x_i x_j.$$

The right side is $\beta\sqrt{\omega/n}$ times an average of the numbers $x_i x_j \in \{-1,1\}$, so each coordinate of ∇f is at most $\beta\sqrt{\omega/n}$ in absolute value. Summing over the $n(n-1)/2$ coordinates, we find $|\nabla f|^2 \le \beta^2\omega(n-1)/2 = \sigma_W^2$; in particular, f is σ_W-Lipschitz. Let $\bar f = \mathbb{E}_{\widetilde{J}}f$. By the second bound in (4.19) (Proposition 4.21), we have $\mathbb{E}_{\widetilde{J}}\exp\{\theta(f-\bar f)\} \le e^{\theta^2\sigma_W^2/2}$ for every $\theta \in \mathbb{R}$. The exponential Markov inequality then implies the two one-sided bounds

$$\mathbb{P}_{\widetilde{J}}(f - \bar f \ge a) \le e^{-a^2/(2\sigma_W^2)}, \qquad \mathbb{P}_{\widetilde{J}}(f - \bar f \le -a) \le e^{-a^2/(2\sigma_W^2)} \qquad (a \ge 0). \tag{10.7}$$

Note that $W \ge \mathbb{E}_{\widetilde{J}}W/2$ if and only if $f \ge \log(\mathbb{E}_{\widetilde{J}}W/2)$. If $\log(\mathbb{E}_{\widetilde{J}}W/2) > \bar f$, then (10.6) and the first bound in (10.7) with $a = \log(\mathbb{E}_{\widetilde{J}}W/2) - \bar f$ give

$$\frac{e^{-\Theta}}{8} \le \mathbb{P}_{\widetilde{J}}\big(f \ge \log(\mathbb{E}_{\widetilde{J}}W/2)\big) \le \exp\Big\{-\frac{(\log(\mathbb{E}_{\widetilde{J}}W/2) - \bar f)^2}{2\sigma_W^2}\Big\},$$

so that $\bar f \ge \log(\mathbb{E}_{\tilde J} W/2) - \sigma_W\sqrt{2(\Theta+\log 8)}$. This lower bound for $\bar f$ also holds if $\log(\mathbb{E}_{\tilde J} W/2) \le \bar f$. The second bound in (10.7) with $a = \sigma_W\sqrt{2\log 2}$ then implies that

$$\mathbb{P}_{\tilde J}\Big(f > \log(\mathbb{E}_{\tilde J} W/2) - \sigma_W\sqrt{2(\Theta+\log 8)} - \sigma_W\sqrt{2\log 2}\Big) \ge \frac{1}{2}.$$

We estimate the loss in the threshold using $\sigma_W^2 \le \beta^2 n\omega/2$, $\sqrt{u+v} \le \sqrt{u} + \sqrt{v}$, and $\sqrt{\log 8} + \sqrt{\log 2} = (1+\sqrt{3})\sqrt{\log 2} < 3$:

$$\sigma_W\sqrt{2\Theta} + \sigma_W\big(\sqrt{2\log 8} + \sqrt{2\log 2}\big) \le \frac{\beta^2 n\omega q_*}{\sqrt{2}} + 3\beta\sqrt{n\omega}.$$

Recalling that $\log(\mathbb{E}_{\tilde J} W/2) = \beta^2\omega(n-1)/4 - \log 2$, we conclude that, with conditional probability at least $1/2$,

$$\log W \ge \frac{\beta^2\omega(n-1)}{4} - \beta^2 n\omega q_* - 3\beta\sqrt{n\omega} - \log 2. \tag{10.8}$$

The constants in (10.8), and the threshold for n beyond which it holds, do not depend on J or S.

Step 2: comparison with the rotation estimate. By the linearity of H_n^J in J, we have $H_n^{J'} = \sqrt{1-\omega}\, H_n^J + \sqrt{\omega}\, H_n^{\tilde J}$. Because $1 - \sqrt{1-\omega} > 0$ and $H_n^J \le ne_1$ on S', we have

$$\frac{Z_n^{J'}}{Z_n^J} = \pi\Big(\exp\big\{-\beta\big(1-\sqrt{1-\omega}\big)H_n^J + \beta\sqrt{\omega}H_n^{\tilde J}\big\}\Big) \ge \pi(S')\, e^{-\beta(1-\sqrt{1-\omega})ne_1}\, W.$$

For $0 \le \omega \le 1$, we have $1 - \sqrt{1-\omega} = \omega/(1+\sqrt{1-\omega}) \le \omega/(2-\omega) \le \omega/2 + \omega^2/2$; the last inequality is equivalent to $\omega(1-\omega) \ge 0$. We take logarithms and insert $\pi(S') \ge e^{-D_1 L}/2$, (10.8), and $\beta^2\omega(n-1)/4 \ge \beta^2 n\omega/4 - \beta^2/8$. Recalling that $n\omega = \Lambda L$, we obtain, with conditional probability at least $1/2$,

$$\log Z_n^{J'} - \log Z_n^J \ge n\omega\Big(\frac{\beta^2}{4} - \frac{\beta e_1}{2} - \beta^2 q_*\Big) - D_1 L - \mathrm{Err}_n = \big[(g_0 - \beta^2 q_*)\Lambda - D_1\big]L - \mathrm{Err}_n,$$

where

$$\mathrm{Err}_n = \frac{\beta e_1 \Lambda^2 L^2}{2n} + 3\beta\sqrt{\Lambda L} + \frac{\beta^2}{8} + 2\log 2.$$

By the choices of q_* and Λ, we have $\beta^2 q_* < g_0/8$ and $D_1 < g_0\Lambda/8 - 1$, so that

$$(g_0 - \beta^2 q_*)\Lambda - D_1 > \frac{7}{8}g_0\Lambda - \frac{1}{8}g_0\Lambda = \frac{3}{4}g_0\Lambda.$$

Since $L \to \infty$ and $L/n \to 0$, we have $\mathrm{Err}_n = o(L)$; this is where the two conditions on the scale are used together. We have $\mathrm{Err}_n \le g_0\Lambda L/4$ for large n, and then

$$\mathbb{P}_{\tilde J}\Big(\log Z_n^{J'} - \log Z_n^J \ge \frac{g_0\Lambda L}{2}\Big) \ge \frac{1}{2} \qquad \text{for every } J \in \mathcal{E}_n \cap \mathcal{V}_n. \tag{10.9}$$

The threshold for n depends only on β, D_1, and the sequence (L_n).

Since the event in (10.9) does not involve S, we need no measurable choice of S as a function of J (in [77, Section 6], the first violating set in a fixed order of the subsets is chosen; either way works). By Fubini's theorem, the left side of (10.9) is a measurable function of J. Integrating it over J, we obtain

$$\frac{1}{2}\mathbb{P}(\mathcal{E}_n \cap \mathcal{V}_n) \le \mathbb{P}_{J,\widetilde{J}}\Big(\log Z_n^{J'} - \log Z_n^{J} \ge \frac{g_0 \Lambda L}{2}\Big).$$

The right side is a probability under the joint law of $(J, \widetilde{J})$, which is the setting of Proposition 4.28. Since $0 < \omega < 1/2$, the last bound of Proposition 4.28, with $t = \omega$ and $u = g_0 \Lambda L/2 = g_0 n\omega/2$, bounds it by $\exp\{-u^2/(2\beta^2 n\omega)\} = \exp\{-g_0^2 \Lambda L/(8\beta^2)\}$. We conclude that

$$\mathbb{P}(\mathcal{V}_n) \le \mathbb{P}(\mathcal{E}_n^c) + \mathbb{P}(\mathcal{E}_n \cap \mathcal{V}_n) \le e^{-c_E n} + 2\exp\Big\{-\frac{g_0^2 \Lambda L}{8\beta^2}\Big\} \longrightarrow 0.$$

This bound holds simultaneously for all subsets, without a union bound over them.

Step 3: from pairs to targets. Let $\mathcal{C}_n(D_1) = \mathcal{V}_n^c$, fix $J \in \mathcal{C}_n(D_1)$ and $S \subseteq \Sigma_n$, and let U_S be the set of $x \in S$ with $\pi(\{z \in S : |R(x,z)| \ge q_*\}) < e^{-D_3 L}$. Suppose that $\pi(U_S) \ge e^{-D_1 L}$. Since $U_S \subseteq S$, every $x \in U_S$ satisfies $\pi(\{y \in U_S : |R(x,y)| \ge q_*\}) < e^{-D_3 L}$. This implies that

$$\begin{aligned}\pi(\cdot \mid U_S)^{\otimes 2}(|R| \ge q_*) &= \frac{1}{\pi(U_S)^2} \sum_{x \in U_S} \pi(x)\, \pi(\{y \in U_S : |R(x,y)| \ge q_*\}) \\ &< \frac{e^{-D_3 L}}{\pi(U_S)} \le e^{-(D_3 - D_1)L} < e^{-D_2 L},\end{aligned}$$

where in the last step we used $D_3 > D_1 + D_2$. Then U_S satisfies (10.3), which contradicts $J \notin \mathcal{V}_n$. We conclude that $\pi(U_S) < e^{-D_1 L}$ for every S, which is (10.1). Finally, for all sufficiently large n, the last display of Step 2 and $L \le n$ give

$$\mathbb{P}\big(\mathcal{C}_n(D_1)^c\big) = \mathbb{P}(\mathcal{V}_n) \le e^{-c_E n} + 2\exp\Big\{-\frac{g_0^2 \Lambda L}{8\beta^2}\Big\} \le 3e^{-D_4 L}, \qquad D_4 = \min\Big\{c_E, \frac{g_0^2 \Lambda}{8\beta^2}\Big\}.$$

The constant D_4 depends only on β and D_1. This completes the proof. □

For $D > 0$, we write $c(D)$ for the constant D_3 given by Lemma 10.1 with $D_1 = D$. It depends only on β and D.

10.2 Parameters and exceptional triples

Exponents

We fix exponents λ and ξ of the scales such that

$$0 < \lambda < \xi < \frac{1}{4}, \qquad 2\xi < 3\lambda, \tag{10.10}$$

and we assume the following upper bound on the mean bias $\Delta_{n,\beta} = P_\beta - p_n(\beta)$: there exists $C_b < \infty$ such that

$$0 \le P_\beta - p_n(\beta) \le C_b n^{-3\lambda} \qquad \text{for all sufficiently large } n. \tag{10.11}$$

For Theorem 1.1, we take $\lambda = 7/30$ and $\xi = 5/21$, and (10.11) is Theorem 4.11, with exponent $7/10 = 3\lambda$. Then (10.10) holds because $7/30 < 5/21 < 1/4$ and $2\xi = 10/21 < 7/10 = 3\lambda$. The barrier exponent is one third of the exponent of the mean-bias input. Table 10.2 lists the powers of n that appear in this section and in Sections 11 and 12; the error terms of the static estimates are tabulated in Section 9.

The fixed parameters

Let $q_* = q_*(\beta)$ be the overlap of Lemma 10.1. Recall from Lemma 9.2 that for each fixed large-overlap cutoff $b_* \in (0,1)$, there exists a small-overlap cutoff $\underline{b}(b_*) > 0$, depending only on β and b_*, such that the three-replica pressure deficit (9.3) holds for all feasible overlap matrices $D(r,q,\delta)$ with $q \ge b_*$, $|r+\delta| \le \underline{b}(b_*)$, $|r-\delta| \le \underline{b}(b_*)$, and $|q - \Gamma(q)| \le \nu_{\mathrm{fix}}(b_*)$. We fix signed-overlap cutoffs b_1 and b_2 with

$$0 < b_1 < 1, \quad b_1 \le \underline{b}(q_*), \qquad 0 < b_2 < b_1, \quad b_2 \le \underline{b}(b_1),$$

and then $t_0 > 0$ with

$$4t_0 < \min\{q_*, b_2, b_1^2/2\}, \qquad 3t_0 < q_\beta. \tag{10.12}$$

In Section 11, the thresholds lie in $[t_0, 2t_0]$, and the cells that we define below use the cutoff $3t_0$. We use the first condition in (10.12) to keep these values below q_* and b_2, with room for the error terms of Section 11. It also gives $3t_0 < b_1^2/2$, which we use there to bound the number of deleted references. The second condition ensures that the intervals J_β and I_β chosen next lie inside $(0, q_\beta)$, where $\Gamma(t) = t$ by (4.10). We fix compact intervals J_β and I_β such that

$$[t_0, 2t_0] \subseteq \operatorname{int} J_\beta, \qquad J_\beta \subseteq \operatorname{int} I_\beta, \qquad I_\beta \subseteq (t_0/2, 3t_0). \tag{10.13}$$

The centers of all three-replica constraints met in Section 11 lie in J_β. The local comparison of Sections 8 and 9 is applied only at centers in J_β, with its local cutoff chosen so small that all its windows lie in I_β. Let $m_\beta > 0$ be the lower bound of (5.20) for $\alpha(r)^2\beta^6 \mathbb{E}\chi(r, Y_r)^3$ on I_β; it is positive because I_β is a compact subset of $(0, q_\beta)$.

We fix the band of the local comparison of Section 7 at $\vartheta = 1/2$ and $E = 1/100$. Here ϑ is the ratio $(q-r)/h$ of that comparison and $E\rho_n \le |\delta| \le 2E\rho_n$ is its range of imbalances. At these values (7.37) holds and $G(\vartheta, E) > 1/3$. We fix the pressure coefficient g, a fraction $\varkappa \in (0,1)$, and the scale constant $R > 0$ with

$$0 < g < m_\beta E^2 G(\vartheta, E), \tag{10.14}$$

$$R^3 > \frac{3C_b}{\varkappa g}, \tag{10.15}$$

where C_b is the constant of (10.11). Here g is the coefficient of ρ_n^3 in the local fractional gap (8.19), and $\varkappa$ is the fraction of this gap that is kept in the probability bounds of Section 9. The scales are

$$\rho_n = Rn^{-\lambda}, \qquad \varepsilon_n = E\rho_n, \qquad \eta_n = n^{-\xi}, \tag{10.16}$$

which are the scales of Remark 9.8. With these choices, the data of the simultaneous estimate in Section 9.3 are the intervals $J_\beta \subseteq \operatorname{int} I_\beta$, the large-overlap cutoffs q_* and b_1, the band (ϑ, E), the coefficient g, and the fraction $\varkappa$. The remaining constants of the static estimates (the local cutoff, the tolerance ν of the pair exclusions, the pressure deficits of the separated-overlap, fixed-scale, and pair regimes, the coefficient of the fractional tilt, and the denominator margin (9.19)) are fixed in Sections 8 and 9 with positive slack. They depend only on β and on the parameters above, and not on n.

Finally, we fix integers

$$H = \lceil 2/b_1^2 \rceil, \qquad B \geq H + 2. \tag{10.17}$$

Here H bounds the number of references deleted at the start of a block in Section 11, and B is the number of levels in a block; B is chosen in Section 12. After B, the constants c_s, C_s of the schedule (10.24) below are fixed, and then the coefficient $A > 0$ of the logarithmic time scale

$$L_n = An^{\lambda}$$

is chosen in Section 12, with $A = A_\beta$ for Theorem 1.1; the time horizon is e^{L_n}. The sequence L_n tends to infinity and is $o(n)$, because $0 < \lambda < 1$; in particular, Lemma 10.1 applies to it.

We note that the quantities that depend on n are the scales ρ_n, ε_n, and η_n of (10.16), the logarithmic time scale L_n, the number N_{lev} of levels (fixed in Section 11), the list sizes K_s of (10.24), and the total number N of list entries. We never let a fixed parameter vary with n, and the threshold τ for the total variation distance does not enter until the last step of Section 12.

Exceptional triples

The overlap constraints that we use in Section 11 are of two kinds. The first kind states that when x and $\widetilde{v}$ have a large overlap, their small overlaps with a third configuration v are nearly equal. For a large-overlap cutoff $b_* \in (0,1)$ and a small-overlap cutoff $b' \in (0,1)$, we define

$$\begin{aligned}\mathcal{B}_n(b_*, b') = \big\{(v, x, \widetilde{v}) \in \Sigma_n^3 : R(x, \widetilde{v}) \geq b_*,\ |R(v,x)| \leq b',\ |R(v,\widetilde{v})| \leq b',\\ |R(v,x) - R(v,\widetilde{v})| > 2\eta_n\big\}.\end{aligned} \tag{10.18}$$

This is the set excluded in [77, Lemma 4.1], with the tolerance $2n^{-1/100}$ there replaced by $2\eta_n$. We bound its Gibbs probability at this tolerance in Section 9 through the pressure

Table 10.1. The fixed parameters of the dynamical argument, in the order in which they are chosen. Every entry is fixed before $n \to \infty$; each depends only on β and on the entries above it.

Parameter	Choice	Reference
q_β	right end of $\operatorname{supp} \mu_\beta = [0, q_\beta]$	Theorem 4.7
q_*	overlap of the overlap-neighborhood mass bound	Lemma 10.1
$b_1,\ b_2$	$b_1 \le \underline{b}(q_*)$, $b_2 \le \underline{b}(b_1)$, $0 < b_2 < b_1 < 1$	(9.3)
t_0	$4t_0 < \min\{q_*, b_2, b_1^2/2\}$, $3t_0 < q_\beta$	(10.12)
$J_\beta,\ I_\beta$	compact intervals around $[t_0, 2t_0]$	(10.13)
m_β	lower bound for the scalar coefficient on I_β	(5.20)
$\vartheta,\ E$	$\vartheta = 1/2$, $E = 1/100$	(7.37)
g	$0 < g < m_\beta E^2 G(\vartheta, E)$	(10.14)
$\varkappa$	$0 < \varkappa < 1$	Section 9
C_b	constant of the mean-bias input; for Theorem 1.1, that of Theorem 4.11	(10.11)
R	$R^3 > 3C_b/(\varkappa g)$	(10.15)
static constants	local cutoff, ν, deficits, fractional tilt, denominator margin	Sections 8 and 9
H	$H = \lceil 2/b_1^2 \rceil$	(10.17)
threshold constants	widths and spacing of the threshold intervals	Section 11
B	integer, $B \ge H + 2$, block length	Section 12
$c_s,\ C_s$	schedule of list sizes, $0 \le s < B$	(10.24)
A	coefficient of $L_n = An^\lambda$	Section 12

estimate (9.3), and not through [77, Lemma 4.1]. A triple $(v, x, \widetilde{v}) \notin \mathcal{B}_n(b_*, b')$ such that $R(x, \widetilde{v}) \ge b_*$, $|R(v, x)| \le b'$, and $|R(v, \widetilde{v})| \le b'$ satisfies $|R(v, x) - R(v, \widetilde{v})| \le 2\eta_n$.

The second kind concerns overlaps near a level r in the compact interval J_β. For $(v, x, \widetilde{v}) \in \Sigma_n^3$, we write

$$r = \frac{R(v, x) + R(v, \widetilde{v})}{2}, \qquad \delta = \frac{R(v, x) - R(v, \widetilde{v})}{2}, \qquad q = R(x, \widetilde{v}), \tag{10.19}$$

so that the Gram matrix of $(v, x, \widetilde{v})$ is the matrix $D(r, q, \delta)$ of Section 9, with off-diagonal entries $r + \delta$, $r - \delta$, and q. Let $\mathcal{N}_n$ be the set of triples with

$$r \in J_\beta, \qquad \varepsilon_n \le |\delta| \le 2\varepsilon_n, \qquad q \ge r + \rho_n.$$

A triple outside $\mathcal{N}_n$ with $r \in J_\beta$ and $\varepsilon_n \le |\delta| \le 2\varepsilon_n$ satisfies $R(x, \widetilde{v}) < r + \rho_n$: if the overlaps of v with x and with $\widetilde{v}$ lie near a common level r and differ by an amount of order ε_n, then the overlap of x and $\widetilde{v}$ cannot exceed r by ρ_n. The set $\mathcal{N}_n$ corresponds to

Table 10.2. Powers of n at the scales (10.16) and the logarithmic time scale $L_n = An^\lambda$, for general exponents and for Theorem 1.1.

Quantity	General	Theorem 1.1	Role
$P_\beta - p_n(\beta)$	$O(n^{-3\lambda})$	$O(n^{-7/10})$	mean-bias input
ρ_n	$Rn^{-\lambda}$	$Rn^{-7/30}$	overlap gap
η_n	$n^{-\xi}$	$n^{-5/21}$	fixed-scale tolerance
L_n	An^λ	$A_\beta n^{7/30}$	logarithmic time scale
$n\rho_n^3$	$R^3 n^{1-3\lambda}$	$R^3 n^{3/10}$	narrow triples
$n\eta_n^2$	$n^{1-2\xi}$	$n^{11/21}$	fixed-scale triples
$n\rho_n^3/L_n$	$(R^3/A)n^{1-4\lambda}$	$(R^3/A_\beta)n^{1/15}$	(10.22)
$n\eta_n^2/L_n$	$n^{1-2\xi-\lambda}/A$	$n^{61/210}/A_\beta$	(10.22)
η_n/ρ_n	$n^{\lambda-\xi}/R$	$n^{-1/210}/R$	Section 11

the set excluded in [77, Lemma 7.1]. There, the gap is $\rho = n^{-1/10000}$, the imbalance is of order ρ^{20}, and the centers are restricted by the condition $\mu_\beta([r-\rho^8, r+\rho^8]) \le \rho^4$ on the Parisi measure. Here the gap ρ_n and the imbalance $\varepsilon_n = E\rho_n$ are of the same order, and there is no condition on the Parisi measure, because the local comparison of Section 8 is uniform over centers in J_β. The exceptional set is

$$\mathcal{B}_n = \mathcal{N}_n \cup \mathcal{B}_n(q_*, b_1) \cup \mathcal{B}_n(b_1, b_2). \tag{10.20}$$

It is a deterministic subset of Σ_n^3, and it is not closed under permutations or spin flips of the coordinates. In the dynamical argument, we evaluate $\mathcal{B}_n$ on triples in several orders and orientations; since $\pi^{\otimes 3}$ is invariant under these operations, each of them costs only a fixed factor in Lemma 10.4.

The Gibbs probability of $\mathcal{B}_n$ is controlled in Section 9, where different parts of it are handled by different estimates. The set $\mathcal{N}_n$ is the set of narrow triples of Section 9.3 for the interval of centers J_β. By Lemma 9.6, a narrow triple either has its pair $(x, \widetilde{v})$ excluded by the pair deficit (9.2), or its Gram matrix lies in the local family, for which we use the positive fractional moment (8.19), or in the separated-overlap family, for which we use the stronger pressure deficit (9.17) and ordinary Gaussian concentration. The sets $\mathcal{B}_n(q_*, b_1)$ and $\mathcal{B}_n(b_1, b_2)$ are contained in the set $\mathcal{W}_n$ of fixed-scale triples of Section 9.3, which is controlled by (9.3) and the pair deficit. We need no fractional moment bound at a distant overlap q.

Lemma 10.2 (One disorder event for the exceptional triples)**.** *Let (ρ_n) and (η_n) be deterministic positive sequences with $\rho_n \to 0$ that satisfy* (S1)–(S5)*, and let $\varepsilon_n = E\rho_n$. Let $\mathcal{B}_n$ be the set* (10.20) *defined with these sequences, and let $\mathcal{A}_n$ be the disorder event of Lemma 9.7 for these sequences and for the data of Section 9.3 fixed in Table 10.1 (in particular the intervals J_β and I_β*

and the large-overlap cutoffs q_* and b_1). Then $\mathbb{P}(\mathcal{A}_n) \to 1$, and there exist constants $C, c > 0$ such that, for all sufficiently large n and every $J \in \mathcal{A}_n$,

$$\pi^{\otimes 3}(\mathcal{B}_n) \le q_n, \qquad q_n = C(n+1)^3\big(e^{-cn\rho_n^3} + e^{-cn\eta_n^2} + e^{-cn}\big). \tag{10.21}$$

Proof. By Lemma 9.7, we have $\mathbb{P}(\mathcal{A}_n^c) \to 0$. Moreover, for large n and every $J \in \mathcal{A}_n$, (9.21) bounds the Gibbs probability of the set $\mathcal{N}_n$ of narrow triples of Section 9.3 by $C(n+1)^3 e^{-cn\rho_n^3}$, and that of the set $\mathcal{W}_n$ of fixed-scale triples of Section 9.3 by $C(n+1)^3 e^{-cn\eta_n^2}$.

It remains to compare $\mathcal{B}_n$ with these two sets, where our triples $(v, x, \widetilde{v})$ are the triples (v, x, y) of Section 9. The set $\mathcal{N}_n$ in (10.20) is the set of narrow triples for the interval of centers J_β. Let $(v, x, \widetilde{v}) \in \mathcal{B}_n(b_*, b')$ with $(b_*, b') = (q_*, b_1)$ or $(b_*, b') = (b_1, b_2)$. Then $R(x, \widetilde{v}) \ge b_*$, $|R(v, x)| \le b' \le \underline{b}(b_*)$, $|R(v, \widetilde{v})| \le b' \le \underline{b}(b_*)$, and $|R(v, x) - R(v, \widetilde{v})| > 2\eta_n$. Since b_* is one of the cutoffs q_* and b_1, the triple satisfies the conditions that define $\mathcal{W}_n$. This shows that $\mathcal{B}_n(q_*, b_1) \cup \mathcal{B}_n(b_1, b_2) \subseteq \mathcal{W}_n$. Summing the two bounds of the previous paragraph, we obtain $\pi^{\otimes 3}(\mathcal{B}_n) \le C(n+1)^3(e^{-cn\rho_n^3} + e^{-cn\eta_n^2})$, which implies (10.21). □

We apply Lemma 10.2 with the scales (10.16). Since $\lambda > 0$, we have $\rho_n = Rn^{-\lambda} \to 0$. These scales also satisfy (S1)–(S5), as in Remark 9.8(b). Indeed, by (10.11) and (10.15), we have $\Delta_{n,\beta}/\rho_n^3 \le C_b/R^3 < \varkappa g/3$ for large n, which is (S1). Further, $n^{7/15}/(n^2\rho_n^6) = R^{-6}n^{6\lambda - 23/15}$ tends to zero because $6\lambda < 3/2 < 23/15$; the quantities $n\rho_n^4 = R^4 n^{1-4\lambda}$ and $n\eta_n^4 = n^{1-4\xi}$ grow faster than $\log n$ because $\lambda, \xi < 1/4$; and $\rho_n^3/\eta_n^2 = R^3 n^{2\xi - 3\lambda}$ tends to zero because $2\xi < 3\lambda$. From now on, $\mathcal{B}_n$, $\mathcal{A}_n$, and q_n denote the objects of Lemma 10.2 at the scales (10.16). In Section 12.7, the lemma is applied with $\eta_n = \rho_n/1000$ instead.

Remark 10.3 (Two kinds of error). In the proof of Theorem 1.1, we use two kinds of small quantities, and we never mix them. The bound (9.20) on the disorder probability $\mathbb{P}(\mathcal{A}_n^c)$ includes a polynomial term from Chebyshev's inequality, and therefore it does not remain small after multiplication by a factor e^{CL_n}. For $\lambda = 7/30$ and $\xi = 5/21$, the terms $e^{-cn\rho_n^4}$ and $e^{-cn\eta_n^4}$ in it, which are $e^{-cR^4 n^{1/15}}$ and $e^{-cn^{1/21}}$, also decay more slowly than $e^{-CL_n} = e^{-CAn^{7/30}}$ for every fixed $C > 0$. We remove this disorder probability, together with the probabilities of the complements of the events of Lemma 10.1, once, before counting reference configurations or path positions. The conditional Gibbs bound q_n in (10.21), in contrast, holds for every $J \in \mathcal{A}_n$, and it absorbs every factor e^{CL_n}. Table 10.2 shows that the exponents $n\rho_n^3$, $n\eta_n^2$, and n exceed L_n by a power of n. We verify in the proof of Lemma 10.4 that

$$\lim_{n\to\infty} e^{CL_n} q_n = 0 \qquad \text{for every fixed } C > 0. \tag{10.22}$$

In Lemma 10.4, only q_n is multiplied by the number of reference configurations. The deterministic set $\mathcal{B}_n$ contains every exceptional triple that can arise for the allowed thresholds and paths, and (10.21) bounds its mass under $\pi^{\otimes 3}$ by q_n. We impose no condition on the disorder that depends on the thresholds of Section 11.

10.3 Fixed families of reference configurations

The induction of Section 11 selects references from lists of independent Gibbs samples, organized in blocks of B consecutive levels. This is the construction of [77, Section 8.3]. The selection cell at a level is determined by the references chosen at the preceding levels of the same block. The list sizes are chosen recursively over the B positions of a block, so that the depth of the recursion is the fixed number B even though the number of levels tends to infinity. In [77, Section 8.3], the required properties of the lists are proved with probability tending to one jointly over the disorder, the lists, and the path. The proof there uses union bounds over path positions and over triples, and the bounds for the triples are averaged over the disorder. At our scales, the averaged bound available from Lemma 9.7 contains the disorder failure probability of Remark 10.3, and this term does not remain small after multiplication by the number $e^{O(L_n)}$ of reference configurations. We therefore fix the disorder first, in the event of Lemma 10.2, and use only the conditional bound (10.21). At such a disorder, we fix one realization of the lists, together with a set V of exceptional targets, before any path is considered.

Lists, cells, and list sizes

Let $N_{\mathrm{lev}} = N_{\mathrm{lev}}(n) \geq 1$ be the number of levels; it is fixed in Section 11. Level $j \in \{0, \dots, N_{\mathrm{lev}} - 1\}$ belongs to block $\lfloor j/B \rfloor$ and has position $s(j) = j - B\lfloor j/B \rfloor$ in its block. The list at level j is an ordered tuple

$$\mathcal{F}_j = (y_{j,1}, \dots, y_{j,K_{s(j)}}) \in \Sigma_n^{K_{s(j)}},$$

whose coordinates we call the entries of $\mathcal{F}_j$; the sizes K_s are defined in (10.24) below. We write

$$\mathcal{F}_j^{\pm} = \{\sigma y : \sigma \in \{-1, 1\},\ y \text{ an entry of } \mathcal{F}_j\}$$

for the signed list. The family of reference configurations is $\mathcal{F} = (\mathcal{F}_0, \dots, \mathcal{F}_{N_{\mathrm{lev}}-1})$, and $N = \sum_{j < N_{\mathrm{lev}}} K_{s(j)}$ is its total number of entries.

For a level j, let $\mathcal{T}_j$ be the set of tuples $T = (u_i)_{i \in \mathcal{I}}$, where $\mathcal{I}$ is a set of levels of the block of j that precede j, and u_i is an entry of $\mathcal{F}_i$ for each $i \in \mathcal{I}$. The cell of T is

$$S_T = \{x \in \Sigma_n : |R(u, x)| \leq 3t_0 \text{ for every entry } u \text{ of } T\}. \tag{10.23}$$

The empty tuple has cell $S_\emptyset = \Sigma_n$, and it is the only element of $\mathcal{T}_j$ when $s(j) = 0$. Spin flips of the entries do not change the cells. The references selected in Section 11 at the levels of a block that precede j, after removing their signs, form a tuple in $\mathcal{T}_j$, and the cells used there are among the cells S_T.

The list sizes are chosen by the following schedule over positions $s = 0, \dots, B-1$. Recall that $c(D)$ denotes the constant D_3 of Lemma 10.1 for $D_1 = D$. We set $C_0 = 0$ and, recursively,

$$c_s = c(C_s + 5), \qquad C_{s+1} = C_s + c_s + 2, \qquad K_s = \lceil e^{(c_s+1)L_n} \rceil. \tag{10.24}$$

We have $C_s = \sum_{s'<s}(c_{s'} + 2)$. The constants c_s and C_s depend only on β and s, and only K_s depends on n and A. For $D > 0$, let $\mathcal{C}_n(D)$ be the event of Lemma 10.1 for the sequence (L_n). On $\mathcal{C}_n(D)$, (10.1) reads

$$\pi\big(\{x \in S : \pi(\{y \in S : |R(x,y)| \geq q_*\}) < e^{-c(D)L_n}\}\big) < e^{-DL_n} \qquad \text{for every } S \subseteq \Sigma_n. \tag{10.25}$$

Only B of these events are needed. We define the disorder event

$$\mathcal{G}_n = \mathcal{A}_n \cap \bigcap_{s=0}^{B-1} \mathcal{C}_n(C_s + 5). \tag{10.26}$$

It is measurable, and $\mathbb{P}(\mathcal{G}_n) \to 1$ by Lemmas 10.1 and 10.2, since it is the intersection of a fixed number of events whose probabilities tend to one.

The family

Lemma 10.4 (Fixed families of reference configurations)**.** *Let $B \geq 1$ be a fixed integer, let (L_n) be a deterministic sequence with $L_n \to \infty$ and $L_n/n \to 0$, and let K_s and $\mathcal{G}_n$ be given by* (10.24) *and* (10.26)*. Let $N_{\mathrm{lev}} = N_{\mathrm{lev}}(n) \geq 1$ be integers with $N_{\mathrm{lev}} \leq e^{L_n}$, and assume that q_n satisfies* (10.22)*. Then, for all sufficiently large n and every $J \in \mathcal{G}_n$, there exist lists $\mathcal{F}_j \in \Sigma_n^{K_{s(j)}}$, $0 \leq j < N_{\mathrm{lev}}$, and a set $V \subseteq \Sigma_n$ with*

$$\pi(V) \leq 3e^{-4L_n}, \tag{10.27}$$

such that the following four properties hold.

(i) Overlap neighborhoods in every cell of the current block. *For every level j, every $T \in \mathcal{T}_j$, and every $x \in S_T \setminus V$, there exists an entry y of $\mathcal{F}_j$ with $y \in S_T$ and $|R(x,y)| \geq q_*$.*

(ii) Comparison across the overlap q_*. *For every $x \in \Sigma_n \setminus V$, all levels $i \neq j$, and all $v \in \mathcal{F}_i^{\pm}$ and $\widetilde{v} \in \mathcal{F}_j^{\pm}$ with*

$$R(x,\widetilde{v}) \geq q_*, \qquad |R(v,x)| \leq b_1, \qquad |R(v,\widetilde{v})| \leq b_1,$$

we have $|R(v,x) - R(v,\widetilde{v})| \leq 2\eta_n$.

(iii) Comparison across the overlap b_1 for triples of entries. *For all pairwise distinct levels i, a, and j, and all $v \in \mathcal{F}_a^{\pm}$, $y \in \mathcal{F}_i^{\pm}$, and $\widetilde{v} \in \mathcal{F}_j^{\pm}$ with*

$$|R(y,\widetilde{v})| \geq b_1, \qquad |R(v,y)| \leq b_2, \qquad |R(v,\widetilde{v})| \leq b_2,$$

we have $|R(v,y) - \sigma R(v,\widetilde{v})| \leq 2\eta_n$, where $\sigma \in \{-1,1\}$ is the sign of $R(y,\widetilde{v})$. In particular, $\big||R(v,y)| - |R(v,\widetilde{v})|\big| \leq 2\eta_n$.

(iv) No exceptional narrow triple. *For every $x \in \Sigma_n \setminus V$, all levels $i \neq j$, and all $v \in \mathcal{F}_i^{\pm}$ and $\widetilde{v} \in \mathcal{F}_j^{\pm}$, we have $(v, x, \widetilde{v}) \notin \mathcal{N}_n$. That is, if the coordinates* (10.19) *of $(v, x, \widetilde{v})$ satisfy $r \in J_\beta$ and $\varepsilon_n \le |\delta| \le 2\varepsilon_n$, then $R(x, \widetilde{v}) < r + \rho_n$.*

If $L_n = An^\lambda$ for a fixed $A > 0$, with λ and ξ as in (10.10)*, then $L_n \to \infty$, $L_n/n \to 0$, and* (10.22) *holds.*

The threshold for n depends on β, on the parameters of Table 10.1, and on the sequences (L_n) and (N_{lev}), but not on J. In Sections 11 and 12, the lemma is applied with $L_n = An^\lambda$ and the number of levels fixed there, which is at most a constant times n^λ.

Proof. We adapt the independent-list construction of [77, Section 8.3] by estimating exceptional Gibbs mass at each fixed disorder in $\mathcal{G}_n$, before considering any path. The proof is an expectation argument over independent samples of the lists. We bound the expected Gibbs mass of the targets missed by the lists, that is, of the configurations x such that, at some level j, some cell S_T containing x contains no entry y of $\mathcal{F}_j$ with $|R(x, y)| \ge q_*$. We also bound the probability that three list entries form an exceptional triple, and the expected Gibbs mass of the configurations that form an exceptional triple with two list entries. We then choose a realization at which the sum of the Gibbs mass of the union of these two sets and the indicator of that event is at most its mean.

Step 1: random lists. Fix n and $J \in \mathcal{G}_n$, and write $L = L_n$. We let the entries $y_{j,k}$, $0 \le j < N_{\mathrm{lev}}$, $1 \le k \le K_{s(j)}$, be independent random configurations with law π, defined on an auxiliary probability space with probability $\mathbb{P}_{\mathcal{F}}$ and expectation $\mathbb{E}_{\mathcal{F}}$. The disorder is fixed throughout, and the space Σ_n^N of realizations is finite.

Step 2: targets missed by the lists. For a level j and $T \in \mathcal{T}_j$, let $U_{j,T}$ be the set of $x \in S_T$ such that no entry y of $\mathcal{F}_j$ satisfies $y \in S_T$ and $|R(x, y)| \ge q_*$, and let $V_1 = \bigcup_j \bigcup_{T \in \mathcal{T}_j} U_{j,T}$. Fix j, let $s = s(j)$, and condition on the lists at the levels of the block of j that precede j. This fixes $\mathcal{T}_j$ and the cells S_T, and leaves the K_s entries of $\mathcal{F}_j$ independent with law π. For $T \in \mathcal{T}_j$ and $x \in S_T$, let $a_T(x) = \pi(\{y \in S_T : |R(x, y)| \ge q_*\})$. Then

$$\mathbb{E}_{\mathcal{F}}\big[\pi(U_{j,T}) \;\big|\; \text{earlier lists}\big] = \sum_{x \in S_T} \pi(x)\big(1 - a_T(x)\big)^{K_s}.$$

We split the sum according to whether $a_T(x) < e^{-c_s L}$. Recall that $J \in \mathcal{C}_n(C_s + 5)$ and $c_s = c(C_s + 5)$. By (10.25) with $S = S_T$, the first group has Gibbs mass less than $e^{-(C_s+5)L}$. For the second group, $(1 - a_T(x))^{K_s} \le \exp\{-a_T(x) K_s\} \le \exp\{-e^{-c_s L} e^{(c_s+1)L}\} = e^{-e^L}$. The number of tuples is

$$|\mathcal{T}_j| \le \prod_{s' < s} \big(K_{s'} + 1\big) \le \prod_{s' < s} e^{(c_{s'}+2)L} = e^{C_s L},$$

where we used $K_{s'} + 1 \le e^{(c_{s'}+1)L} + 2 \le e^{(c_{s'}+2)L}$, valid once $e^L \ge 3$. Summing over $T \in \mathcal{T}_j$ and taking expectations, we obtain

$$\mathbb{E}_{\mathcal{F}}\, \pi\Big(\bigcup_{T \in \mathcal{T}_j} U_{j,T} \Big) \le e^{C_s L} \big(e^{-(C_s+5)L} + e^{-e^L} \big) \le 2e^{-5L},$$

where the last inequality holds once $e^L \ge (C_s + 5)L$ for every $s < B$. Summing over the $N_{\mathrm{lev}} \le e^L$ levels, we find

$$\mathbb{E}_{\mathcal{F}}\, \pi(V_1) \le 2e^{-4L}. \tag{10.28}$$

Step 3: exceptional triples. Let $c_{\max} = \max_{s<B} c_s$. Using $N_{\mathrm{lev}} \le e^L$ and $K_s \le e^{(c_s+2)L}$, we find $N \le e^{C_N L}$ with $C_N = c_{\max} + 3$.

Let $\mathcal{O}$ be the event that there exist entries y_1, y_2, and y_3 of lists at three distinct levels, and signs $\sigma_1, \sigma_2, \sigma_3 \in \{-1, 1\}$, such that $(\sigma_1 y_1, \sigma_2 y_2, \sigma_3 y_3) \in \mathcal{B}_n$. Fix a choice of three entries, specified by their levels and indices, and of signs. The three entries are independent with law π, because they belong to distinct levels, and π is invariant under spin flips. The triple therefore has law $\pi^{\otimes 3}$, and it lies in $\mathcal{B}_n$ with probability at most q_n by (10.21). There are at most N^3 ordered choices of three entries and 8 choices of signs, which gives

$$\mathbb{P}_{\mathcal{F}}(\mathcal{O}) \le 8N^3 q_n.$$

The configurations themselves need not be distinct; only their sampling indices are.

Let V_2 be the set of $x \in \Sigma_n$ for which there exist entries y and y' of lists at two distinct levels, signs $\sigma, \sigma', \sigma'' \in \{-1, 1\}$, and an ordering of σy, $\sigma' y'$, and $\sigma'' x$ as a triple, such that this triple lies in $\mathcal{B}_n$. Let x be a configuration with law π, independent of the lists. For a fixed ordered pair of entries at distinct levels, fixed signs, and a fixed ordering, the resulting triple has law $\pi^{\otimes 3}$, and it lies in $\mathcal{B}_n$ with probability at most q_n. There are at most N^2 ordered pairs of entries, 8 choices of signs, and 6 orderings, and a union bound yields

$$\mathbb{E}_{\mathcal{F}}\, \pi(V_2) = \mathbb{P}_{\mathcal{F} \otimes \pi}(x \in V_2) \le 48N^2 q_n.$$

Step 4: choice of one realization. Let $V = V_1 \cup V_2$. Combining (10.28) with the two bounds of Step 3, we conclude that

$$\mathbb{E}_{\mathcal{F}} \big[\mathbf{1}_{\mathcal{O}} + \pi(V) \big] \le 2e^{-4L} + \big(8N^3 + 48N^2 \big) q_n \le 2e^{-4L} + 56 e^{3C_N L} q_n \le 3e^{-4L},$$

where the last inequality holds for all sufficiently large n by (10.22) with $C = 3C_N + 4$. On the finite probability space of Step 1, some realization of the lists satisfies $\mathbf{1}_{\mathcal{O}} + \pi(V) \le \mathbb{E}_{\mathcal{F}}[\mathbf{1}_{\mathcal{O}} + \pi(V)] \le 3e^{-4L}$. We fix it. Then $\pi(V) \le 3e^{-4L}$, which is (10.27), and the realization does not belong to $\mathcal{O}$, since $\mathbf{1}_{\mathcal{O}} \in \{0, 1\}$ and $3e^{-4L} < 1$.

Step 5: the four properties. Property (i) holds because $V \supseteq V_1$. For (ii), let x, v, and $\widetilde{v}$ be as there. The triple $(v, x, \widetilde{v})$ is one of the orderings in the definition of V_2, with $\sigma'' = 1$. Since $x \notin V_2$, it does not lie in $\mathcal{B}_n \supseteq \mathcal{B}_n(q_*, b_1)$. By the definition (10.18) and the hypotheses of (ii), this means that $|R(v,x) - R(v,\widetilde{v})| \le 2\eta_n$. The same argument with $\mathcal{N}_n \subseteq \mathcal{B}_n$ proves (iv). For (iii), let v, y, $\widetilde{v}$, and σ be as there. Then $R(y, \sigma\widetilde{v}) = |R(y,\widetilde{v})| \ge b_1$, $|R(v,y)| \le b_2$, and $|R(v,\sigma\widetilde{v})| \le b_2$. Since the realization does not belong to $\mathcal{O}$, the triple $(v, y, \sigma\widetilde{v})$ of signed entries from three distinct levels does not lie in $\mathcal{B}_n(b_1, b_2)$. This implies that $|R(v,y) - \sigma R(v,\widetilde{v})| \le 2\eta_n$, and the last claim of (iii) follows from $\big||R(v,y)| - |R(v,\widetilde{v})|\big| \le |R(v,y) - \sigma R(v,\widetilde{v})|$.

Step 6: the time scale $L_n = An^\lambda$. If $L_n = An^\lambda$ with $0 < \lambda < 1$, then $L_n \to \infty$ and $L_n/n \to 0$. For a fixed $C > 0$, we have

$$e^{CL_n} q_n = C'(n+1)^3 \Big(e^{CAn^\lambda - cR^3 n^{1-3\lambda}} + e^{CAn^\lambda - cn^{1-2\xi}} + e^{CAn^\lambda - cn} \Big),$$

with C' the constant of (10.21). By (10.10), we have $1 - 3\lambda > \lambda$ and $1 - 2\xi > 1/2 > \lambda$. Hence there exists $\lambda' > 0$ such that each of the three exponents is at most $-n^{\lambda'}$ for all large n, and $(n+1)^3 e^{-n^{\lambda'}} \to 0$. This proves (10.22) and completes the proof. □

The proof gives slightly more than (ii)–(iv). For every $x \notin V$, no ordering of x or $-x$ with two signed entries from distinct levels lies in $\mathcal{B}_n$, and no ordering of three signed entries from distinct levels lies in $\mathcal{B}_n$. Properties (ii)–(iv) are the instances of these statements used in Section 11.

Remark 10.5 (The same family for every path avoiding V). The family $\mathcal{F}$ and the set V of Lemma 10.4 depend only on J, and properties (i)–(iv) are deterministic statements about $\mathcal{F}$, V, and configurations outside V. No path enters their construction. We may therefore apply them along any finite sequence $x_0, \dots, x_m$ in $\Sigma_n \setminus V$ and for any choice of the thresholds of Section 11: properties (i), (ii), and (iv) hold at every x_k, simultaneously for all k, all levels, and all entries, and (iii) holds for all entries. This is the form in which we use the lemma in the induction of Section 11.

The order of choices matters for the stationary sign test of Section 12. There, we bound the probability that the stationary discrete heat-bath chain visits V at one of the indices $0, \dots, m$ by $(m+1)\pi(V)$, using only that its law is π at each fixed integer index. This union bound requires that V is fixed before the chain is run. Note also that a reference selected along a path depends on the path and need not have law π; for this reason we state (ii)–(iv) for all entries of all lists, and not only for selected references. Finally, we choose the family separately at each $J \in \mathcal{G}_n$ and assert no measurable dependence on J. None is needed, because the quantities that we integrate over the disorder in Section 12 do not depend on the family: they are the Gibbs masses of initial states whose transition law is close to π in total variation.

SECTION 11

Stable overlap constraints along a trajectory

This section contains the deterministic part of the dynamical argument. We fix the disorder, the families of reference configurations constructed in Section 10, an array of thresholds, and a sequence of configurations with small overlap increments that avoids the exceptional set V of Section 10. From these data, an explicit algorithm (Definition 11.2) selects one reference at each level and finds one crossing time at each level, searching backward from the previous crossing. The main result is Proposition 11.10: if the overlap of the sequence with its first reference becomes nonpositive, then every crossing exists, and the errors in all overlap constraints stay of order ε_n. Corollary 11.11 extracts the consequence used in Section 12: at one common time, the overlap of the sequence with the first reference of every block lies within a fixed multiple of ε_n of the corresponding threshold. In Section 12, the thresholds are made random, and we show by a counting bound that this cannot happen before the time in Theorem 1.1.

We use the four properties of the families of reference configurations proved in Lemma 10.4, which we recall as (P1)–(P4) in Section 11.1, and elementary geometry of overlaps. No probability is used in this section.

The construction is that of [77, Sections 8.3 and 8.4]. There, references are selected from fixed lists of Gibbs samples inside cells determined by the earlier references of the current block. The cells are reset to the whole space at the start of each block, references with a large overlap with the new reference are deleted, and each crossing time is found by a backward search from the previous one. It is also proved in [77, Section 8.4] that the deleted references form a final segment, with at most H elements, of the retained references ordered by level, and that older blocks are protected by the complete block that precedes each reset; we give these arguments in Lemma 11.6. What is new is the form of the induction hypothesis. In [77, (8.19)–(8.20)], after j levels, the overlaps of every retained reference with the current target and with the later references are controlled within $6j\varepsilon$, where $\varepsilon = \rho^{20}$ is the scale of the narrow overlap constraints there. These errors accumulate over the levels, and the final count of [77, (8.29)] requires $N_{\mathrm{lev}}\varepsilon$ to be much smaller than ρ for N_{lev} levels. Here we compare each retained reference with the next retained reference. This comparison changes only when a deletion removes the next retained reference, which happens at most once for each reference. The error of every constraint stays bounded by a fixed multiple of ε at every level (Definition 11.4). The argument uses that the first two levels of every block are never deleted, which follows

from the same protection argument. As a result, $\varepsilon_n = E\rho_n$ with a fixed $E > 0$ is allowed.

11.1 Standing assumptions and the threshold geometry

Throughout this section and Section 12, $\beta > 1$ is fixed, n is large, and the disorder J is fixed; we write $\pi = \pi^J$. We use the parameters of Section 10, which are listed in Table 10.1. The overlap $q_* \in (0,1)$ is that of the overlap-neighborhood mass bound, Lemma 10.1. The signed-overlap cutoffs $0 < b_2 < b_1 < 1$ and the level $t_0 > 0$ satisfy (10.12), and in particular

$$4t_0 < \min\{q_*, b_2, b_1^2/2\}.$$

This implies that $3t_0 < b_1^2/2 < b_1$, since $b_1 < 1$. The compact interval J_β of (10.13) contains $[t_0, 2t_0]$ in its interior. The integer $H = \lceil 2/b_1^2 \rceil$ is that of (10.17), and $B \geq H + 2$ is a fixed integer, the number of levels in a block; it is chosen in Section 12. The scales of (10.16) are

$$\rho = \rho_n = Rn^{-\lambda}, \qquad \varepsilon = \varepsilon_n = E\rho_n, \qquad \eta_n = n^{-\xi},$$

where $R > 0$ and $E = 1/100$ are fixed, and the exponents satisfy $0 < \lambda < \xi < 1/4$ by (10.10). Theorem 1.1 uses $\lambda = 7/30$ and $\xi = 5/21$. In Section 12.7 we use $\eta_n = \rho_n/1000$ in place of $\eta_n = n^{-\xi}$ with $\xi > \lambda$; this is why the hypotheses below allow a positive limit of η_n/ρ_n. We suppress n from ρ_n and ε_n, and we abbreviate

$$\zeta = 2\eta_n,$$

which is the tolerance of the overlap comparisons in Lemma 10.4. For the deterministic induction in this section, the asymptotic scale assumptions needed are $\varepsilon = E\rho$ and

$$\rho \to 0, \qquad \frac{\eta_n}{\rho} \to E', \qquad \frac{1}{n\rho} \to 0, \tag{11.1}$$

where $E' \in [0, E/10]$ is fixed. For Theorem 1.1, these hold with $E' = 0$ because $0 < \lambda < \xi$ and $\lambda < 1$; at $\lambda = 7/30$ and $\xi = 5/21$, we have $\eta_n/\rho_n = n^{-1/210}/R$. In Section 12.7, they hold with $E' = E/10 = 1/1000$. The counting estimate in Section 12 also retains $\rho_n = Rn^{-\lambda}$; the additional static scale conditions for Section 12.7 are checked there.

Reference families. The levels are the integers $0 \leq i < N_{\mathrm{lev}}$, where $N_{\mathrm{lev}} = MB$ and M is defined in (11.3) below. As in Section 10.3, level i belongs to the block $\ell(i) = \lfloor i/B \rfloor$; we say that it is the first level of its block if $i = B\ell(i)$. A family of reference configurations consists of an ordered list $\mathcal{F}_i$ of configurations for each level i, whose coordinates are called the entries of $\mathcal{F}_i$, together with a set $V \subseteq \Sigma_n$. We call a configuration σy, where y is an entry of $\mathcal{F}_i$ and $\sigma \in \{-1, 1\}$, a signed entry of level i; the set of signed entries of level i is $\mathcal{F}_i^{\pm}$. For a finite tuple T of configurations, the cell S_T is the set of $x \in \Sigma_n$ with

$|R(y,x)| \le 3t_0$ for every y in T, as in (10.23), and $S_T = \Sigma_n$ for the empty tuple. Reversing the sign of an element of T does not change S_T.

By Lemma 10.4, applied with the logarithmic time scale $L_n = An^\lambda$ of Section 12 and with $N_{\mathrm{lev}} \le e^{L_n}$, for all sufficiently large n and every disorder in the event $\mathcal{G}_n$ of (10.26), there exists a family with $\pi(V) \le 3e^{-4L_n}$ and with the following four properties. They are consequences of properties (i)–(iv) of Lemma 10.4, as we explain after the list.

(P1) *Overlap neighborhoods in every cell.* Let j be a level, let T be a tuple consisting of one signed entry of each earlier level of the block of j, and let $x \in S_T \setminus V$. Then there exists an entry y of $\mathcal{F}_j$ with $y \in S_T$ and $|R(x,y)| \ge q_*$.

(P2) *Comparison across the overlap q_*.* Let $x \notin V$, and let v and $\widetilde{v}$ be signed entries of two distinct levels. If $R(x,\widetilde{v}) \ge q_*$, $|R(v,x)| \le b_1$, and $|R(v,\widetilde{v})| \le b_1$, then $|R(v,x) - R(v,\widetilde{v})| \le \zeta$.

(P3) *Comparison across the overlap b_1.* Let v, y, and $\widetilde{v}$ be signed entries of three distinct levels. If $|R(y,\widetilde{v})| \ge b_1$, $|R(v,y)| \le b_2$, and $|R(v,\widetilde{v})| \le b_2$, then $\big||R(v,y)| - |R(v,\widetilde{v})|\big| \le \zeta$.

(P4) *No narrow triples.* Let $x \notin V$, and let v and $\widetilde{v}$ be signed entries of two distinct levels. Let $r = (R(v,x) + R(v,\widetilde{v}))/2$ and $\delta = (R(v,x) - R(v,\widetilde{v}))/2$. Then it is not the case that

$$r \in J_\beta, \qquad \varepsilon \le |\delta| \le 2\varepsilon, \qquad R(\widetilde{v},x) \ge r + \rho;$$

that is, the triple $(v,x,\widetilde{v})$ does not belong to the set $\mathcal{N}_n$ of narrow triples defined after (10.19).

Property (P1) is the case of property (i) in which the tuple has one entry at each earlier level of the block of j; here we used that the cell of a tuple of signed entries equals the cell of the corresponding tuple of entries. Properties (P2) and (P4) are properties (ii) and (iv), and (P3) is the last assertion of property (iii). Properties (P2) and (P3) say that, across a large overlap between x and $\widetilde{v}$ (or between y and $\widetilde{v}$), a small overlap with v is transferred from one configuration to the other with error at most ζ. The family and V do not depend on any path or threshold, and the four properties are deterministic statements about them (see Remark 10.5).

Paths. We call a finite sequence $x_0, \dots, x_m$ in Σ_n a path if

$$|R(y,x_{k+1}) - R(y,x_k)| \le 2/n \qquad \text{for all } y \in \Sigma_n \text{ and } 0 \le k < m. \tag{11.2}$$

We say that a path avoids V if $x_k \notin V$ for $0 \le k \le m$. If x_{k+1} and x_k differ in at most one coordinate, then (11.2) holds; in particular, every trajectory of the discrete heat-bath chain is a path.

Thresholds. We set

$$M = \Big\lfloor \frac{t_0}{3B\rho} \Big\rfloor, \qquad N_{\mathrm{lev}} = MB, \qquad I_i = \big[t_0 + 3i\rho,\ t_0 + (3i+1)\rho\big] \quad (0 \le i < N_{\mathrm{lev}}). \tag{11.3}$$

For all sufficiently large n, we have $M \geq 1$. Each interval I_i has length ρ and lies in $[t_0, 2t_0]$, since $t_0 + (3(N_{\mathrm{lev}} - 1) + 1)\rho \leq t_0 + 3MB\rho \leq 2t_0$. We call a vector $\mathbf{r} = (r_0, \dots, r_{N_{\mathrm{lev}}-1})$ such that $r_i \in I_i$ for every i a threshold array. For every threshold array and all levels $i < a$, we have

$$r_a - r_i \geq \big(3(a-i) - 1\big)\rho \geq 2\rho. \tag{11.4}$$

Radii and margins. We define

$$\begin{aligned} &\psi_0 = \zeta + \frac{2}{n}, && \psi_1 = 2\varepsilon + 2\zeta + \frac{2}{n}, \\ &\Psi_0 = \psi_0 + 2\varepsilon = 2\varepsilon + \zeta + \frac{2}{n}, && \Psi_1 = \psi_1 + 2\varepsilon = 4\varepsilon + 2\zeta + \frac{2}{n}, \qquad \Psi_2 = \Psi_1 + \zeta. \end{aligned} \tag{11.5}$$

In the induction, ψ_0 and ψ_1 bound the error, relative to the threshold, of the overlap of a retained reference with its next retained reference; Ψ_0 and Ψ_1 bound the corresponding error for the overlap with the current target; and Ψ_2 bounds it for the overlaps between retained references. We use the following inequalities:

$$\begin{gathered} \zeta < 2\varepsilon, \qquad \frac{1}{n} \leq \varepsilon, \qquad 2t_0 + \Psi_2 < \min\{3t_0, b_2\}, \qquad 2t_0 + \Psi_1 < q_*, \\ 2\Psi_2 + \zeta < 2\rho, \qquad 5\varepsilon + 3\zeta + \frac{7}{n} \leq \rho, \end{gathered} \tag{11.6}$$

and the closed neighborhood of $[t_0, 2t_0]$ of radius $3\varepsilon + 2\zeta + 3/n$ is contained in J_β.

Lemma 11.1 (The margins hold for large n)**.** *Assume* (11.1)*. There exists n_0, depending only on the fixed parameters q_*, b_2, t_0, J_β, R, E, E', and the sequences (ρ_n) and (η_n), such that* (11.6) *and the inclusion of the neighborhood in J_β hold for all $n \geq n_0$.*

Proof. By (11.5), we have

$$\frac{\zeta}{\rho} \to 2E', \qquad \frac{2\Psi_2 + \zeta}{\rho} = 8E + \frac{7\zeta}{\rho} + \frac{4}{n\rho} \to 8E + 14E', \qquad \frac{5\varepsilon + 3\zeta + 7/n}{\rho} \to 5E + 6E',$$

where we used (11.1) and $\varepsilon = E\rho$. Since $E' \leq E/10$ and $E = 1/100$, the three limits are at most $E/5 < 2E$, $9.4E < 2$, and $5.6E < 1$, respectively. The inequalities in the second line of (11.6) hold for large n, and so do $\zeta < 2\varepsilon$ and $1/n \leq \varepsilon$. The radii Ψ_2 and Ψ_1 tend to zero, while $2t_0 < 3t_0$, $2t_0 < b_2$, and $2t_0 < q_*$ by (10.12). This gives the remaining two inequalities. Finally, the compact set $[t_0, 2t_0]$ lies in the interior of J_β and has positive distance from its complement, while the radius $3\varepsilon + 2\zeta + 3/n$ tends to zero. □

We record three consequences of (11.6) that are used repeatedly. First, we have $\Psi_1 < \Psi_2 < t_0$, and every overlap within Ψ_2 of a threshold $r_i \in [t_0, 2t_0]$ is positive and has absolute value less than $2t_0 + \Psi_2 < 3t_0$. Second, such an overlap has absolute value less than $b_2 < b_1$, because $2t_0 + \Psi_2 < b_2$. Third, every threshold is smaller than q_*, because $r_i \leq 2t_0 < q_*$.

11.2 The selection and crossing algorithm

We now define the references and the crossing times. The algorithm is that of [77, Section 8.4], including its conventions on failure, with zero-based levels.

Definition 11.2 (Selection and crossing algorithm)**.** Fix a family of reference configurations, a path $x_0, \dots, x_m$, and a threshold array $\mathbf{r}$.

(i) *Level* 0. If some entry y of $\mathcal{F}_0$ satisfies $|R(y, x_0)| \geq q_*$, let y be the first such entry and set $v_0 = \sigma y$, where σ is the sign of $R(y, x_0)$, so that $R(v_0, x_0) \geq q_*$. If the set $\{0 \leq k \leq m : R(v_0, x_k) \leq r_0\}$ is also nonempty, let k_0 be its smallest element and set $\mathcal{R}_0 = \{0\}$; level 0 then succeeds. If no entry qualifies, let v_0 be the first entry of $\mathcal{F}_0$; level 0 then fails, as it does when v_0 is selected but the set defining k_0 is empty.

(ii) *Level $j \geq 1$, when levels $0, \dots, j-1$ have succeeded.* Write $\bar{k} = k_{j-1}$, let T_j be the tuple (v_l) over the levels $l < j$ of the block of j (empty if j is the first level of its block), and write $S_j = S_{T_j}$.

  (a) *Selection.* If some entry y of $\mathcal{F}_j$ satisfies $y \in S_j$ and $|R(y, x_{\bar{k}})| \geq q_*$, let y be the first such entry and set $v_j = \sigma y$, where σ is the sign of $R(y, x_{\bar{k}})$, so that $R(v_j, x_{\bar{k}}) \geq q_*$. Otherwise let v_j be the first entry of $\mathcal{F}_j$; level j then fails.

  (b) *Deletion.* The conflicts at level j are the elements of $\mathcal{D}_j = \{i \in \mathcal{R}_{j-1} : |R(v_i, v_j)| > b_1\}$, and we set $\mathcal{R}_j = (\mathcal{R}_{j-1} \setminus \mathcal{D}_j) \cup \{j\}$. The elements of $\mathcal{R}_{j-1} \setminus \mathcal{D}_j$ are called compatible.

  (c) *Crossing.* If the set $\{0 \leq k \leq \bar{k} : R(v_j, x_k) \leq r_j\}$ is nonempty, let k_j be its largest element; level j then succeeds. Otherwise level j fails, and v_j is kept as the reference of level j.

(iii) *After a failure.* If some level fails, the algorithm stops, and for every later level i we let v_i be the first entry of $\mathcal{F}_i$, without a change of sign.

We say that the algorithm succeeds through level j if levels $0, \dots, j$ all succeed. We write $v(x_0)$ for the reference v_0 of level 0.

The reference $v(x_0)$ depends only on x_0 and $\mathcal{F}_0$. The set $\mathcal{R}_j$ is the set of retained levels after level j, and the references v_i with $i \in \mathcal{R}_j$ are the retained references. We order $\mathcal{R}_j$ increasingly; its largest element is j, the level of the newest reference. The crossing time at level 0 is the first time k with $R(v_0, x_k) \leq r_0$, and the crossing time at level $j \geq 1$ is the last time $k \leq k_{j-1}$ with $R(v_j, x_k) \leq r_j$; in particular $k_0 \geq k_1 \geq \cdots$. These crossing times are not stopping times of the path, and nothing below requires them to be.

The convention in (ii)(c), that a selected reference whose crossing fails is kept, matters for the following lemma. If v_j were replaced by a default after its crossing failed, it would depend on its own threshold r_j.

Lemma 11.3 (References depend only on earlier thresholds)**.** *Fix the family and the path. For every level i, the reference v_i is a function of $(r_0, \dots, r_{i-1})$ alone, with finitely many values, each taken on a Borel set; the reference $v_0 = v(x_0)$ does not depend on* $\mathbf{r}$*.*

Proof. Let ok_j be the indicator that the algorithm succeeds through level j. We show by induction on j that v_j is a finite-valued Borel function of $(r_0, \dots, r_{j-1})$, that ok_j is a Borel function of $(r_0, \dots, r_j)$, and that k_j and $\mathcal{R}_j$, restricted to $\{\mathrm{ok}_j = 1\}$, are finite-valued Borel functions of $(r_0, \dots, r_j)$. For $j = 0$, the selection uses only x_0 and $\mathcal{F}_0$, and v_0 is constant. If no entry qualifies, then $\mathrm{ok}_0 = 0$ identically. Otherwise v_0 is fixed, and the event $\{\mathrm{ok}_0 = 1, k_0 = k\}$ is described by the finitely many inequalities $R(v_0, x_l) > r_0$ for $l < k$ and $R(v_0, x_k) \le r_0$; on it, $\mathcal{R}_0 = \{0\}$. For $j \ge 1$, on $\{\mathrm{ok}_{j-1} = 0\}$ the reference v_j is the fixed default and $\mathrm{ok}_j = 0$. On $\{\mathrm{ok}_{j-1} = 1\}$, the cell S_j, the target $x_{k_{j-1}}$, and therefore the selection, depend only on $v_0, \dots, v_{j-1}$ and k_{j-1}, which are Borel functions of $(r_0, \dots, r_{j-1})$ there by the induction hypothesis. The threshold r_j enters only through the crossing, which does not change v_j, and for fixed earlier data the event $\{\mathrm{ok}_j = 1, k_j = k\}$ is described by finitely many inequalities between r_j and the overlaps $R(v_j, x_l)$. On $\{\mathrm{ok}_j = 1\}$, the set $\mathcal{R}_j$ is determined by $\mathcal{R}_{j-1}$ and $v_0, \dots, v_j$. □

Next retained references. Suppose that the algorithm succeeds through level $j-1$ and that the selection at level j succeeds, so that $\mathcal{R}_j$ is defined (for $j = 0$, suppose that level 0 succeeds). For $i \in \mathcal{R}_j \setminus \{j\}$, the index of the next retained reference of i after level j is

$$a_j(i) = \min\{a \in \mathcal{R}_j : a > i\},$$

the next retained reference of i is $v_{a_j(i)}$, and we write $c_j(i) = R(v_i, v_{a_j(i)})$. We say that the next retained reference of i is unchanged if $a_j(i) = i + 1$, and replaced otherwise. In Definition 11.4 below, the overlap $c_j(i)$ is the center around which the overlaps of v_i with the configurations of the path are controlled.

Definition 11.4 (The invariant after level j)**.** Suppose that the algorithm succeeds through level j. The invariant $\mathbf{I}(j)$ consists of the following five statements.

(a) *Blocks.* Every level of the block of j that is at most j belongs to $\mathcal{R}_j$. For every block $\ell < \ell(j)$, the set of levels of block ℓ that belong to $\mathcal{R}_j$ is $\{\ell B, \ell B + 1, \dots, \ell B + B - d - 1\}$ for some $0 \le d \le H$.

(b) *The newest reference.* We have

$$R(v_j, x_{k_j}) \in [r_j - 2/n,\, r_j]. \tag{11.7}$$

(c) *Centers.* For every $i \in \mathcal{R}_j \setminus \{j\}$, we have $|c_j(i) - r_i| \le \psi_0$ if the next retained reference of i is unchanged. Otherwise $|c_j(i) - r_i| \le \psi_1$, and $a_j(i) = B(\ell(i) + 1)$ is the first level of the block that follows the block of i.

(d) *Windows.* For every $i \in \mathcal{R}_j \setminus \{j\}$, we have

$$|R(v_i, x_{k_j}) - c_j(i)| < 2\varepsilon. \tag{11.8}$$

(e) *Pairs.* For all $i < a$ in $\mathcal{R}_j$, we have

$$|R(v_i, v_a) - r_i| \le \Psi_2. \tag{11.9}$$

Statement (a) implies that $0 \in \mathcal{R}_j$: level 0 is the first level of block 0, and since $B - H \ge 2$, the first two levels of every earlier block $\ell < \ell(j)$ are retained. Combining (c) and (d), we find that every $i \in \mathcal{R}_j \setminus \{j\}$ satisfies

$$|R(v_i, x_{k_j}) - r_i| < \begin{cases} \Psi_0 & \text{if the next retained reference of } i \text{ is unchanged,} \\ \Psi_1 & \text{if it is replaced.} \end{cases} \tag{11.10}$$

11.3 One level of the algorithm

The following five lemmas carry out one step of the induction. They share these hypotheses, which we call (H_j): the inequalities (11.6) and the inclusion of the neighborhood in J_β hold; the family satisfies (P1)–(P4); the path $x_0, \dots, x_m$ avoids V; $\mathbf{r}$ is a threshold array; $1 \le j < N_{\mathrm{lev}}$; the algorithm succeeds through level $j-1$; and $\mathbf{I}(j-1)$ holds. We write $\bar{k} = k_{j-1}$ for the previous crossing time and call $x_{\bar{k}}$ the target of level j. Statements (a), (c), and (e) of $\mathbf{I}(j)$ concern only the set $\mathcal{R}_j$ and the references; in the lemmas below we use them as soon as $\mathcal{R}_j$ is defined, before the crossing at level j is known to exist.

Lemma 11.5 (Selection and transfer)**.** *Assume* (H_j).

(a) For every $i \in \mathcal{R}_{j-1}$, we have $|R(v_i, x_{\bar{k}}) - r_i| \le \Psi_1$. More precisely, the error is at most $2/n$ if $i = j-1$, and less than Ψ_0 if $i < j-1$ and the next retained reference of i after level $j-1$ is unchanged.

(b) The selection at level j succeeds, and $R(v_j, x_{\bar{k}}) \ge q_*$.

(c) For every compatible $i \in \mathcal{R}_{j-1}$, we have

$$|R(v_i, v_j) - R(v_i, x_{\bar{k}})| \le \zeta, \qquad |R(v_i, v_j) - r_i| \le \Psi_2. \tag{11.11}$$

Proof. Part (a) is (11.7) at level $j-1$ for $i = j-1$, and (11.10) at level $j-1$ otherwise; note that $2/n \le \Psi_0 < \Psi_1$.

For part (b), if the tuple T_j is nonempty, then j is not the first level of its block, the block of j is the block of $j-1$, and every level l in T_j belongs to $\mathcal{R}_{j-1}$ by $\mathbf{I}(j-1)$(a). Using part (a) and (11.6), we find $|R(v_l, x_{\bar{k}})| \le r_l + \Psi_1 \le 2t_0 + \Psi_1 < 3t_0$ for every such l. Therefore $x_{\bar{k}} \in S_j$. Since also $x_{\bar{k}} \notin V$, property (P1), applied with the tuple T_j, gives an entry y of $\mathcal{F}_j$ with $y \in S_j$ and $|R(y, x_{\bar{k}})| \ge q_*$, and the selection succeeds.

For part (c), let $i \in \mathcal{R}_{j-1}$ be compatible, which means that $|R(v_i, v_j)| \le b_1$. By part (a) and (11.6), we have $|R(v_i, x_{\bar{k}})| \le 2t_0 + \Psi_1 < b_1$. Further, $R(x_{\bar{k}}, v_j) \ge q_*$ by part (b), and v_i and v_j are signed entries of the distinct levels $i < j$. Property (P2), applied with $x = x_{\bar{k}}$, $v = v_i$, and $\widetilde{v} = v_j$, implies the first inequality in (11.11). The second follows from the first and part (a), since $\zeta + \Psi_1 = \Psi_2$. □

Lemma 11.6 (Deletion and protection of initial segments)**.** *Assume* (H_j).

(a) The set $\mathcal{D}_j$ of conflicts is a final segment of $\mathcal{R}_{j-1}$: if $i < a$ belong to $\mathcal{R}_{j-1}$ and $i \in \mathcal{D}_j$, then $a \in \mathcal{D}_j$.

(b) We have $|\mathcal{D}_j| < 2/b_1^2 \le H$.

(c) If j is not the first level of its block, then $\mathcal{D}_j = \emptyset$.

(d) If j is the first level of its block, then $\mathcal{D}_j$ consists of the last d levels of the block $\ell(j) - 1$, for some $0 \le d \le H$. In particular, the first $B - H \ge 2$ levels of the block $\ell(j) - 1$ belong to $\mathcal{R}_j$, and no level of a block $\ell < \ell(j) - 1$ is deleted.

As a consequence, $\mathbf{I}(j)$*(a) holds.*

Proof. The final-segment, conflict-count, and block-protection arguments are those of [77, Section 8.4, "Conflicts occupy a bounded final segment"], applied here with the bounds of $\mathbf{I}(j-1)$. For part (a), suppose to the contrary that $i < a$ belong to $\mathcal{R}_{j-1}$, that $i \in \mathcal{D}_j$ (that is, $|R(v_i, v_j)| > b_1$), and that a is compatible. By (11.9) at level $j-1$, we have $|R(v_i, v_a) - r_i| \le \Psi_2$, and by (11.11), we have $|R(v_a, v_j) - r_a| \le \Psi_2$. By the consequences of (11.6) recorded after Lemma 11.1, both overlaps are positive and have absolute value less than b_2. Since v_a, v_i, and v_j are signed entries of the three distinct levels a, i, and j, property (P3), applied with $v = v_a$, $y = v_i$, and $\widetilde{v} = v_j$, gives $|R(v_a, v_j)| - |R(v_a, v_i)| \le \zeta$. On the other hand, using (11.4) and (11.6), we have

$$|R(v_a, v_j)| - |R(v_a, v_i)| = R(v_a, v_j) - R(v_i, v_a) \ge (r_a - \Psi_2) - (r_i + \Psi_2) \ge 2\rho - 2\Psi_2 > \zeta,$$

which is a contradiction.

For part (b), we may assume that $d = |\mathcal{D}_j| \ge 1$. For $i \in \mathcal{D}_j$, let σ_i be the sign of $R(v_i, v_j)$, and consider the unit vectors $\mathbf{e} = v_j/\sqrt{n}$ and $\mathbf{e}_i = \sigma_i v_i/\sqrt{n}$ in $\mathbb{R}^n$. Then $\langle \mathbf{e}_i, \mathbf{e} \rangle = |R(v_i, v_j)| > b_1$, and for distinct $i, i' \in \mathcal{D}_j$, (11.9) at level $j-1$ and the consequences of (11.6) recorded after Lemma 11.1 give $|\langle \mathbf{e}_i, \mathbf{e}_{i'} \rangle| = |R(v_i, v_{i'})| < 3t_0$. By the Cauchy–Schwarz inequality,

$$d^2 b_1^2 < \Big\langle \sum_{i \in \mathcal{D}_j} \mathbf{e}_i, \mathbf{e} \Big\rangle^2 \le \Big\| \sum_{i \in \mathcal{D}_j} \mathbf{e}_i \Big\|^2 = d + \sum_{i \ne i'} \langle \mathbf{e}_i, \mathbf{e}_{i'} \rangle \le d + 3t_0 d(d-1).$$

Dividing by d, we obtain $d(b_1^2 - 3t_0) < 1 - 3t_0$. Since $3t_0 < b_1^2/2$, this implies

$$d < \frac{1 - 3t_0}{b_1^2 - 3t_0} < \frac{2}{b_1^2} \le H,$$

where the middle inequality is equivalent to $3t_0(2 - b_1^2) < b_1^2$.

For part (c), the level $j-1$ belongs to the block of j. Then v_{j-1} is an element of T_j, and since $v_j \in S_j$, we have $|R(v_{j-1}, v_j)| \leq 3t_0 < b_1$. In other words, the largest element $j-1$ of $\mathcal{R}_{j-1}$ is compatible, and the final segment $\mathcal{D}_j$ is empty by part (a).

For part (d), let $j = \ell B$ with $\ell \geq 1$. The level $j-1$ is the last level of block $\ell - 1$, which is the block of $j-1$. By $\mathbf{I}(j-1)$(a), all B levels of block $\ell - 1$ belong to $\mathcal{R}_{j-1}$, and they are its B largest elements. By parts (a) and (b), $\mathcal{D}_j$ is a final segment of $\mathcal{R}_{j-1}$ with fewer than $H < B$ elements. It therefore consists of the last $d = |\mathcal{D}_j| \leq H$ levels of block $\ell - 1$.

Finally, we verify $\mathbf{I}(j)$(a). If j is not the first level of its block, nothing is deleted by part (c), and $\mathcal{R}_j = \mathcal{R}_{j-1} \cup \{j\}$; statement (a) at level $j-1$ gives statement (a) at level j. If j is the first level of its block, the block of j meets $\mathcal{R}_j$ in $\{j\}$, the retained levels of block $\ell(j) - 1$ form an initial segment of length $B - d$ with $d \leq H$ by part (d), and the earlier blocks are unchanged. □

Part (d) is the protection of initial segments: once the reset at the first level of block $\ell + 1$ has occurred, the retained levels of block ℓ form an initial segment of length at least $B - H \geq 2$ that is never changed afterward. Indeed, a later deletion removes only levels of the block immediately preceding the current reset, and that block is complete when the reset occurs. This is proved by induction along the run of the algorithm, using only $\mathbf{I}(j-1)$.

Lemma 11.7 (Next retained references after a deletion)**.** *Assume* (H_j)*, so that* $\mathcal{R}_j$ *is defined by Lemma 11.5(b).*

(a) Let $i \in \mathcal{R}_j \setminus \{j\}$ *with* $i \neq j-1$*. If* $a_j(i) \neq a_{j-1}(i)$*, then* j *is the first level of its block,* $\mathcal{D}_j \neq \emptyset$*,* i *is the largest element of* $\mathcal{R}_{j-1} \setminus \mathcal{D}_j$ *and lies in block* $\ell(j) - 1$*,* $a_{j-1}(i) = i + 1$*, and* $a_j(i) = j$*.*

(b) For every $i \in \mathcal{R}_j \setminus \{j\}$*, the bounds of* $\mathbf{I}(j)$*(c) hold, and* $|R(v_i, x_{\bar{k}}) - c_j(i)| < 2\varepsilon$*.*

(c) The statement $\mathbf{I}(j)$*(e) holds.*

Proof. For part (a), note that i is not the largest element $j-1$ of $\mathcal{R}_{j-1}$, and that $a_{j-1}(i)$ is defined. Since $\mathcal{R}_j$ is obtained from $\mathcal{R}_{j-1}$ by removing $\mathcal{D}_j$ and adding j, which exceeds every element of $\mathcal{R}_{j-1}$, the next retained reference of i changes only if $a_{j-1}(i) \in \mathcal{D}_j$. Then $\mathcal{D}_j \neq \emptyset$, and by Lemma 11.6(c), j is the first level of its block. Since $\mathcal{D}_j$ is a final segment of $\mathcal{R}_{j-1}$ by Lemma 11.6(a), every element of $\mathcal{R}_{j-1}$ larger than i lies in $\mathcal{D}_j$. This means that i is the largest element of $\mathcal{R}_{j-1} \setminus \mathcal{D}_j$ and that $a_j(i) = j$. By Lemma 11.6(d), the first $B - H$ levels of block $\ell(j) - 1$ survive, and i lies in this block. Since this block was complete in $\mathcal{R}_{j-1}$, we have $a_{j-1}(i) = i + 1$.

For part (b), there are three cases. *Case 1:* $i \neq j-1$ *and* $a_j(i) = a_{j-1}(i)$. Then $c_j(i) = c_{j-1}(i)$, the next retained reference of i is unchanged after level j if and only if it is unchanged after level $j-1$, and the claims are $\mathbf{I}(j-1)$(c) and (d). *Case 2:* $i = j-1$.

Then $a_j(i) = j = i+1$, and the next retained reference of i is unchanged. Using (11.11) and (11.7) at level $j-1$, we find

$$|c_j(i) - R(v_i, x_{\bar{k}})| \le \zeta, \qquad |c_j(i) - r_i| \le \zeta + \frac{2}{n} = \psi_0,$$

and $|R(v_i, x_{\bar{k}}) - c_j(i)| \le \zeta < 2\varepsilon$ by (11.6). *Case 3: $i \neq j-1$ and $a_j(i) \neq a_{j-1}(i)$.* By part (a), $a_{j-1}(i) = i+1$ and $a_j(i) = j = B(\ell(i)+1)$, which is the first level of the block that follows the block of i. Since $i+1 \in \mathcal{D}_j$, we have $j \neq i+1$, and the next retained reference of i is replaced. Before the deletion, we have $|c_{j-1}(i) - r_i| \le \psi_0$ and $|R(v_i, x_{\bar{k}}) - c_{j-1}(i)| < 2\varepsilon$ by $\mathbf{I}(j-1)$(c) and (d). By (11.11), $|c_j(i) - R(v_i, x_{\bar{k}})| \le \zeta$. Combining these three bounds, we obtain

$$|c_j(i) - r_i| < \zeta + 2\varepsilon + \psi_0 = \psi_1,$$

and $|R(v_i, x_{\bar{k}}) - c_j(i)| \le \zeta < 2\varepsilon$.

For part (c), a pair $i < a$ of elements of $\mathcal{R}_j$ with $a < j$ is a pair of elements of $\mathcal{R}_{j-1}$, and (11.9) at level $j-1$ applies. If $a = j$, then i is compatible, and (11.11) gives $|R(v_i, v_j) - r_i| \le \Psi_2$. □

The next lemma treats the backward search for the crossing at level j. Its main point is that the windows of all old references are preserved simultaneously. To show this, we take the largest index at which some window fails, for all references together. At that index every window is still almost preserved, and in particular so is the window of the next retained reference of the failing reference. This gives the lower bound on the overlap q in part (b) of the lemma.

Lemma 11.8 (The backward search)**.** *Assume* (H_j). *Let $\underline{k} = k_j$ if the set defining k_j in Definition 11.2(ii)(c) is nonempty, and $\underline{k} = 0$ otherwise.*

(a) For every k with $\underline{k} \le k \le \bar{k}$, we have $R(v_j, x_k) \ge r_j - 2/n$. If the crossing exists, then $k_j < \bar{k}$.

(b) Suppose that $|R(v_i, x_k) - c_j(i)| \ge 2\varepsilon$ for some $k \in [\underline{k}, \bar{k}]$ and some $i \in \mathcal{R}_j \setminus \{j\}$. Let k be the largest such index, let i be such that the inequality holds at k, write $a = a_j(i)$, and set

$$r = \frac{R(v_i, x_k) + c_j(i)}{2}, \qquad \delta = \frac{R(v_i, x_k) - c_j(i)}{2}, \qquad q = R(v_a, x_k).$$

Then

$$\varepsilon \le |\delta| \le 2\varepsilon, \qquad |r - r_i| \le 3\varepsilon + 2\zeta + \frac{3}{n}, \qquad q - r \ge 2\rho - 5\varepsilon - 3\zeta - \frac{7}{n} \ge \rho, \quad (11.12)$$

so that (v_i, x_k, v_a) is a narrow triple in the sense of (P4).

(c) For every k with $\underline{k} \le k \le \bar{k}$ and every $i \in \mathcal{R}_j \setminus \{j\}$, we have $|R(v_i, x_k) - c_j(i)| < 2\varepsilon$.

Proof. For part (a), suppose first that the crossing exists. By the definition of k_j as a largest element, we have $R(v_j, x_k) > r_j$ for $k_j < k \le \bar{k}$. By Lemma 11.5(b), we have $R(v_j, x_{\bar{k}}) \ge q_* > r_j$, which implies that $k_j < \bar{k}$. Then (11.2) gives $R(v_j, x_{k_j}) \ge R(v_j, x_{k_j+1}) - 2/n > r_j - 2/n$. If the crossing does not exist, then $R(v_j, x_k) > r_j$ for every $0 \le k \le \bar{k}$.

For part (b), the windows hold at $k = \bar{k}$ by Lemma 11.7(b). This implies that $k < \bar{k}$ and $k+1 \in [\underline{k}, \bar{k}]$. By the maximality of k, every window holds strictly at $k+1$, and (11.2) gives

$$|R(v_{i'}, x_k) - c_j(i')| < 2\varepsilon + \frac{2}{n} \qquad \text{for every } i' \in \mathcal{R}_j \setminus \{j\}. \tag{11.13}$$

For the chosen i, this means that $2\varepsilon \le 2|\delta| < 2\varepsilon + 2/n$, and (11.6) implies that $\varepsilon \le |\delta| < \varepsilon + 1/n \le 2\varepsilon$. We have $|r - r_i| \le |c_j(i) - r_i| + |\delta|$, and the bounds on the centers in $\mathbf{I}(j)$(c), which hold by Lemma 11.7(b), give

$$|r - r_i| < \begin{cases} \psi_0 + \varepsilon + 1/n = \varepsilon + \zeta + 3/n & \text{if } a = i+1, \\ \psi_1 + \varepsilon + 1/n = 3\varepsilon + 2\zeta + 3/n & \text{otherwise.} \end{cases}$$

We now bound q from below, distinguishing three cases. *Case 1: $a = j$.* By part (a), $q = R(v_j, x_k) \ge r_j - 2/n$. *Case 2: $a < j$ and $a = i+1$.* Then $a \in \mathcal{R}_j \setminus \{j\}$. Using (11.13) for $i' = a$ and $|c_j(a) - r_a| \le \psi_1$, we find

$$q > c_j(a) - 2\varepsilon - \frac{2}{n} \ge r_a - \psi_1 - 2\varepsilon - \frac{2}{n} = r_a - 4\varepsilon - 2\zeta - \frac{4}{n}.$$

Case 3: $a < j$ and $a \ne i+1$. By $\mathbf{I}(j)$(c), a is the first level of a block. The level $a+1$ belongs to the same block and satisfies $a+1 \le j$. If this block is the block of j, then $a+1 \in \mathcal{R}_j$ by $\mathbf{I}(j)$(a); otherwise $a+1 \in \mathcal{R}_j$ because the retained initial segment of the block has length at least $B - H \ge 2$. The next retained reference of a is then unchanged, and $|c_j(a) - r_a| \le \psi_0$. By (11.13) for $i' = a$, we have

$$q > r_a - \psi_0 - 2\varepsilon - \frac{2}{n} = r_a - 2\varepsilon - \zeta - \frac{4}{n}.$$

In Case 2 we use the first bound on $|r - r_i|$, and in Cases 1 and 3 the second. In all three cases, we obtain

$$q - r \ge r_a - r_i - 5\varepsilon - 3\zeta - \frac{7}{n} \ge 2\rho - 5\varepsilon - 3\zeta - \frac{7}{n} \ge \rho,$$

where we used (11.4) and the last inequality in (11.6). This proves (11.12). Further, $r_i \in [t_0, 2t_0]$, and the bound on $|r - r_i|$ together with the inclusion of the neighborhood of $[t_0, 2t_0]$ in J_β implies that $r \in J_\beta$. Finally, $R(v_i, x_k) = r + \delta$, $R(v_i, v_a) = c_j(i) = r - \delta$, and $R(v_a, x_k) = q \ge r + \rho$, which are the conditions in (P4) for the triple (v_i, x_k, v_a).

For part (c), suppose that the conclusion fails. Then part (b) produces a narrow triple (v_i, x_k, v_a) in which $x_k \notin V$, because the path avoids V, and v_i and v_a are signed entries of the distinct levels $i < a$. This contradicts (P4). □

Lemma 11.9 (The crossing exists)**.** *Assume* (H_j)*. Then the crossing at level* j *exists, the algorithm succeeds through level* j*, and* (11.7) *holds at level* j*.*

Proof. Suppose that the set defining k_j is empty, so that $\underline{k} = 0$ in Lemma 11.8. The level 0 belongs to $\mathcal{R}_j$ by Lemma 11.6, and $0 \neq j$. By Lemma 11.8(c) at $k = 0$ and by $\mathbf{I}(j)$(c),

$$R(v_0, x_0) < c_j(0) + 2\varepsilon \le r_0 + \psi_1 + 2\varepsilon = r_0 + \Psi_1 \le 2t_0 + \Psi_1 < q_*,$$

where the last inequality is (11.6). Recall that the selection of v_0 at level 0 gives $R(v_0, x_0) \ge q_*$. This contradiction shows that the crossing exists. By Lemma 11.8(a), $R(v_j, x_{k_j}) \ge r_j - 2/n$, and $R(v_j, x_{k_j}) \le r_j$ by the definition of k_j. □

11.4 The induction

Proposition 11.10 (Overlap constraints along a path)**.** *Assume that* (11.6) *and the inclusion of the neighborhood in* J_β *hold, and that the family of reference configurations satisfies (P1)–(P4). Let* $x_0, \dots, x_m$ *be a path that avoids* V*, let* $\mathbf{r}$ *be a threshold array, and run the algorithm of Definition 11.2. If*

$$R\big(v(x_0), x_m\big) \le 0,$$

then the algorithm succeeds through level $N_{\mathrm{lev}} - 1$*, and* $\mathbf{I}(j)$ *holds for every* $0 \le j < N_{\mathrm{lev}}$*. Moreover, the next retained reference of every reference is replaced at most once: if* $i \in \mathcal{R}_j \setminus \{j\}$ *and* $a_j(i) \neq i + 1$*, then* $i \in \mathcal{R}_{j'}$ *and* $a_{j'}(i) = a_j(i)$ *for every* j' *with* $j \le j' < N_{\mathrm{lev}}$*.*

Proof. We first treat level 0. Since $x_0 \notin V$, property (P1), applied with $j = 0$ and the empty tuple, gives an entry y of $\mathcal{F}_0$ with $|R(y, x_0)| \ge q_*$. Then $v_0 = v(x_0)$ is selected, and $R(v_0, x_0) \ge q_* > r_0$. Since $R(v_0, x_m) \le 0 < r_0$, the set defining k_0 is nonempty, and $k_0 \ge 1$. The minimality of k_0 gives $R(v_0, x_{k_0 - 1}) > r_0$, and (11.2) then implies that $R(v_0, x_{k_0}) \in [r_0 - 2/n, r_0]$. Level 0 succeeds, and $\mathbf{I}(0)$ holds: $\mathcal{R}_0 = \{0\}$, statement (b) was just shown, and statements (c)–(e) are empty.

Suppose that $1 \le j < N_{\mathrm{lev}}$, that the algorithm succeeds through level $j - 1$, and that $\mathbf{I}(j-1)$ holds; these are the hypotheses (H_j). By Lemma 11.5(b), the selection at level j succeeds, and by Lemma 11.9, so does the crossing, with (11.7) at level j. Statement $\mathbf{I}(j)$(a) holds by Lemma 11.6, statements (c) and (e) hold by Lemma 11.7, and statement (d) is Lemma 11.8(c) at $k = k_j$, which lies in $[\underline{k}, \bar{k}]$ with $\underline{k} = k_j$. This completes the induction.

For the last assertion, let $a = a_j(i) \neq i + 1$. By $\mathbf{I}(j)$(c), $a = B(\ell(i) + 1)$ is the first level of the block that follows the block of i, and $a \le j$ since $a \in \mathcal{R}_j$. Let $j' > j$. If j' is not the first level of its block, no deletion occurs at level j' by Lemma 11.6(c). Otherwise j' and a are distinct first levels of blocks with $j' > a$, so that $\ell(j') - 1 \ge \ell(a) > \ell(i)$. By Lemma 11.6(d), a deletion at level j' can remove only levels of the block $\ell(j') - 1$ other than its first $B - H$ levels. It follows that i is never deleted, and a, which is the first level

of its block, is never deleted either. The levels strictly between i and a do not belong to $\mathcal{R}_j$, and they never belong to a later set $\mathcal{R}_{j'}$, since the levels added after level j are larger than j. We conclude that $a_{j'}(i) = a$ for every $j' \geq j$. □

The windows of the replaced references are needed inside the induction, because the corresponding references carry constraints through later backward searches; for the count in Section 12, only the first level of each block is used.

Corollary 11.11 (A common time with M narrow windows)**.** *Under the hypotheses of Proposition 11.10, let* $k_* = k_{N_{\mathrm{lev}}-1}$. *Then* $0 \leq k_* \leq m$ *and*

$$|R(v_{\ell B}, x_{k_*}) - r_{\ell B}| < \Psi_0 \qquad \text{for every } \ell \in \{0, 1, \ldots, M-1\}.$$

Proof. Let $j = N_{\mathrm{lev}} - 1 = MB - 1$, the last level of block $M-1$. Let $\ell \leq M-1$. If $\ell < M-1$, then by $\mathbf{I}(j)$(a) the retained levels of block ℓ form an initial segment of length at least $B - H \geq 2$, so that ℓB and $\ell B + 1$ belong to $\mathcal{R}_j$. If $\ell = M-1$, then all levels of block $M-1$ belong to $\mathcal{R}_j$, and $\ell B + 1 \leq j$ since $B \geq 2$. In both cases $\ell B \in \mathcal{R}_j \setminus \{j\}$ and $a_j(\ell B) = \ell B + 1$; that is, the next retained reference of ℓB is unchanged. By statements (c) and (d) of $\mathbf{I}(j)$, we conclude that $|R(v_{\ell B}, x_{k_*}) - r_{\ell B}| < \psi_0 + 2\varepsilon = \Psi_0$. □

SECTION 12

Proof of the barrier theorem

In this section, we prove Theorem 1.1 and Corollary 1.2, and, in Section 12.7, the quantitative and almost-sure forms of Theorem 12.6. We first make the thresholds of Section 11 random and show, using Corollary 11.11, that a path avoiding V keeps a positive overlap with its first reference during the first $T_n = \lceil n^2 e^L \rceil$ update attempts (Proposition 12.2 and Lemma 12.3). This conclusion is deterministic: it holds for every path that avoids V, and the thresholds are only a device in its proof. Combining it with stationarity, spin-flip symmetry, and a Poisson clock, we obtain bounds on the Gibbs mass of the initial states from which the dynamics is close to equilibrium (Proposition 12.5). Integration over the disorder event of Section 10 then proves Theorem 1.1.

The main inputs are Lemma 11.3, Corollary 11.11, the families of reference configurations of Section 10 with the bound $\pi(V) \le 3e^{-4L}$ (see (10.27)), and the properties of the heat-bath dynamics in Lemma 3.2. The probability of the disorder event of Section 10 enters only once, in the last step of the proof of Theorem 1.1. The quantities that are multiplied by the number of time indices are the Gibbs mass $\pi(V)$ and probabilities in an auxiliary space of thresholds, both at a fixed disorder.

The argument follows [77, Sections 8.1 and 8.5]. The product bound for predictable thresholds is the elimination argument following [77, (8.28)]. The reduction of total variation to a sign passage, with the half-space $\{y : R(v, y) \le 0\}$ of Gibbs mass at least $1/2$ and a Poisson clock, is [77, (8.1)–(8.3)], together with the remark on discrete time at the end of [77, Section 8.5]. There are three changes. First, the windows of Corollary 11.11 have radius of order ε and sit at deterministic positions. We therefore count only the M windows at the first levels of the blocks, in place of the union over subsets of size N_{lev}/B in [77, (8.29)]. Second, the static estimates of Section 9 are uniform over the compact interval J_β, and we need no random event restricting the thresholds, such as [77, (8.12)]. As a result, the sign obstruction holds for every path avoiding V, and not only with high probability. Third, the budget T_n of update attempts has logarithm $L + O(\log n)$, in place of $2L$, and every fixed $\tau < 1/2$ is allowed, in place of $\tau = 1/4$.

We keep the notation and the standing assumptions of Section 11.1. Recall that $\rho = Rn^{-\lambda}$, $\varepsilon = E\rho$ with $E = 1/100$, $\zeta = 2\eta_n$, the integers M and $N_{\mathrm{lev}} = MB$ and the intervals I_i are given by (11.3), and the radius $\Psi_0 = 2\varepsilon + \zeta + 2/n$ is given by (11.5).

12.1 Predictable thresholds

Fix the disorder, a family of reference configurations, and a path $x_0, \dots, x_m$. On the product $I_0 \times \cdots \times I_{N_{\mathrm{lev}}-1}$, let $\mathbb{P}^{\mathrm{thr}}$ be the product of the uniform probability measures on the intervals. Under $\mathbb{P}^{\mathrm{thr}}$, the coordinates $r_0, \dots, r_{N_{\mathrm{lev}}-1}$ of a threshold array are independent and r_i is uniform on I_i. This is an auxiliary probability space; it is not related to the randomness of the dynamics. By Lemma 11.3, for every fixed time $k \le m$ the overlap

$$z_i(k) = R(v_i, x_k)$$

of the reference of level i with x_k is a Borel function of $(r_0, \dots, r_{i-1})$. The next lemma is the elimination argument of [77, Section 8.5].

Lemma 12.1 (Windows with predictable centers)**.** *Let $I_0, \dots, I_{\bar N-1}$ be intervals of length $h > 0$, and let $r_0, \dots, r_{\bar N-1}$ be independent random variables with r_i uniform on I_i. Let $D \subseteq \{0, \dots, \bar N - 1\}$, and for $i \in D$ let $a_i \ge 0$ and let z_i be a Borel function of $(r_0, \dots, r_{i-1})$ (a constant if $i = 0$). Then*

$$\mathbb{P}\big\{|z_i - r_i| \le a_i \text{ for every } i \in D\big\} \le \prod_{i \in D} \frac{2a_i}{h}.$$

Proof. We argue by induction on the number of elements of D; the case $D = \varnothing$ is trivial. Let i_* be the largest element of D, let W be the event $\{|z_i - r_i| \le a_i \text{ for every } i \in D \setminus \{i_*\}\}$, and let $\mathscr{H}$ be the σ-algebra generated by $r_0, \dots, r_{i_*-1}$. Then W and z_{i_*} are $\mathscr{H}$-measurable, and r_{i_*} is independent of $\mathscr{H}$. It follows that

$$\mathbb{P}\big\{|z_{i_*} - r_{i_*}| \le a_{i_*} \;\big|\; \mathscr{H}\big\} = \frac{\big|[z_{i_*} - a_{i_*}, z_{i_*} + a_{i_*}] \cap I_{i_*}\big|}{h} \le \frac{2a_{i_*}}{h},$$

where $|\cdot|$ denotes length. Multiplying by $\mathbf{1}_W$ and taking expectations, we obtain

$$\mathbb{P}\big(W \cap \{|z_{i_*} - r_{i_*}| \le a_{i_*}\}\big) \le \frac{2a_{i_*}}{h}\, \mathbb{P}(W),$$

and the induction hypothesis applied to $D \setminus \{i_*\}$ completes the proof. □

The order of integration matters: the largest index is removed first, because the later centers may depend on the earlier thresholds. In the lemma, we condition only on earlier thresholds. We do not condition on the event that the algorithm succeeds, nor on the sets $\mathcal{R}_j$ of retained levels, which depend on all thresholds. A bound $2a_i/h$ larger than one is allowed.

12.2 A deterministic sign obstruction

We define

$$\gamma_n = \frac{2\Psi_0}{\rho} = 4E + \frac{2\zeta}{\rho} + \frac{4}{n\rho}, \tag{12.1}$$

which is the ratio of the length of a window of radius Ψ_0 to the length ρ of a threshold interval. By (11.1), we have $\gamma_n \to 4(E+E') \le 11/250$, and $\gamma_n \to 4E = 1/25$ when $E' = 0$. For integers $m \ge 0$ we set

$$C_n(m) = (m+1)\gamma_n^M. \tag{12.2}$$

Proposition 12.2 (Sign obstruction for paths avoiding V)**.** *Assume that* (11.6) *and the inclusion of the neighborhood in J_β hold, and that the family of reference configurations satisfies (P1)–(P4). Let $m \ge 0$ be an integer with $C_n(m) < 1$. Then*

$$R\big(v(x_0), x_m\big) > 0 \qquad \textit{for every path } x_0, \dots, x_m \textit{ that avoids } V. \tag{12.3}$$

Proof. Fix a path $x_0, \dots, x_m$ that avoids V, and suppose that $R(v(x_0), x_m) \le 0$. Run the algorithm of Definition 11.2 with an arbitrary threshold array $\mathbf{r}$. Recall that the reference $v(x_0)$ of level 0 does not depend on $\mathbf{r}$. The hypothesis of Proposition 11.10 therefore holds for every threshold array. By Corollary 11.11, every threshold array belongs to the union over $0 \le k \le m$ of the events

$$W_k = \big\{\mathbf{r} : |z_{\ell B}(k) - r_{\ell B}| \le \Psi_0 \text{ for every } \ell \in \{0, \dots, M-1\}\big\},$$

where $z_i(k) = R(v_i, x_k)$ and $v_i = v_i(r_0, \dots, r_{i-1})$ is the reference of level i produced by the algorithm. By Lemma 11.3, each $z_{\ell B}(k)$ is a Borel function of $(r_0, \dots, r_{\ell B - 1})$. Applying Lemma 12.1 under $\mathbb{P}^{\mathrm{thr}}$, with $h = \rho$, $D = \{0, B, \dots, (M-1)B\}$, and $a_i = \Psi_0$, we obtain $\mathbb{P}^{\mathrm{thr}}(W_k) \le \gamma_n^M$. Summing over k, we find

$$1 = \mathbb{P}^{\mathrm{thr}}\Big(\bigcup_{k=0}^{m} W_k\Big) \le (m+1)\gamma_n^M = C_n(m) < 1,$$

which is a contradiction. □

The events W_k use deterministic positions and deterministic radii. Their union contains the event produced by the algorithm, but we do not assert that the algorithm produces any particular configuration of retained references, and we do not condition on such an assertion. The proof also uses no property of the crossing times other than those established in Section 11; in particular, no distributional statement about the path at a crossing time is involved.

12.3 The choice of the coefficient

The count $C_n(m)$ is small when $\log(m+1)$ is smaller than $M\log(1/\gamma_n)$. Since $\gamma_n \to 4(E+E') < 1$ and M is of order n^λ, this allows $\log m$ of order n^λ. Fix a coefficient A with

$$0 < A < \frac{t_0 \log\bigl(1/(4E+4E')\bigr)}{3BR}; \tag{12.4}$$

when $E' = 0$, as in the proof of Theorem 1.1, the right side is $t_0 \log 25/(3BR)$. We set

$$L = L_n = An^\lambda, \qquad T_n = \bigl\lceil n^2 e^L \bigr\rceil. \tag{12.5}$$

Here L_n is the logarithmic time scale and e^{L_n} is the time horizon. The integer T_n is a budget of update attempts: the continuous-time dynamics makes on average ne^{L_n} attempts up to time e^{L_n}, and T_n exceeds this number by a factor n.

Lemma 12.3 (The count for paths of exponential length)**.** *Assume* (11.1)*, and let A satisfy* (12.4)*. There exists n_0, depending only on A, t_0, B, R, E, E', and the sequences (ρ_n) and (η_n), such that $C_n(m) < 1$ for every $n \ge n_0$ and every integer $0 \le m \le n^3 e^L$. In particular, this holds for every $0 \le m \le T_n$.*

Proof. By (12.4), there exists $\gamma' \in (4E+4E', 1)$ with $A < t_0 \log(1/\gamma')/(3BR)$; we fix such a γ'. Since $\gamma_n \to 4E+4E'$, we have $\gamma_n \le \gamma'$ for all sufficiently large n. By (11.3), we have $M \ge t_0/(3B\rho) - 1 = t_0 n^\lambda/(3BR) - 1$. Since $C_n(m)$ is increasing in m, it suffices to consider the largest integer m with $m \le n^3 e^L$, for which $m+1 \le 2n^3 e^L$. We find

$$\begin{aligned} \log C_n(m) &\le \log 2 + 3\log n + An^\lambda - \Bigl(\frac{t_0 n^\lambda}{3BR} - 1\Bigr)\log\frac{1}{\gamma'} \\ &= -\Bigl(\frac{t_0 \log(1/\gamma')}{3BR} - A\Bigr)n^\lambda + 3\log n + \log 2 + \log\frac{1}{\gamma'}. \end{aligned}$$

The coefficient of n^λ is negative and $\lambda > 0$, so the right side tends to $-\infty$. Finally, $T_n \le n^2 e^L + 1 \le n^3 e^L$ because $n \ge 2$. □

We make no attempt to optimize the admissible coefficient in (12.4); only its positivity matters for Theorem 1.1.

12.4 Stationarity, the clocks, and total variation

The following lemma turns a sign obstruction for paths into a bound on the Gibbs mass of the initial states from which the dynamics is close to equilibrium. It abstracts the sign-passage reduction of [77, Section 8.1, (8.1)–(8.3)]. It is stated for a general kernel and uses only stationarity at deterministic times, spin-flip symmetry, and the Poisson representation of the continuous-time semigroup; for the latter representation, see [68, Section 20.1, (20.5)–(20.6)].

Lemma 12.4 (From a sign obstruction to total variation)**.** *Let Q be a Markov kernel on Σ_n with stationary law π, and assume that $\pi(-x) = \pi(x)$ for every $x \in \Sigma_n$. Let $v\colon \Sigma_n \to \Sigma_n$ be a map, let $V \subseteq \Sigma_n$, and let $m \geq 0$ be an integer. Assume that every sequence $x_0, \dots, x_k$ with $k \leq m$, with $Q(x_{a-1}, x_a) > 0$ for $1 \leq a \leq k$, and with $x_a \notin V$ for $0 \leq a \leq k$, satisfies $R(v(x_0), x_k) > 0$. Then for every $\tau \in [0, 1/2)$ and every integer $0 \leq k \leq m$,*

$$\pi\big\{x : \|Q^k(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \tau\big\} \leq \frac{(k+1)\pi(V)}{1/2 - \tau}. \tag{12.6}$$

Moreover, for every $t \geq 0$, the kernel $P_t = e^{nt(Q-I)}$ satisfies

$$\pi\big\{x : \|P_t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \tau\big\} \leq \frac{(m+1)\pi(V) + \mathbb{P}\{\mathsf{N} > m\}}{1/2 - \tau}, \tag{12.7}$$

where N has the Poisson law with mean nt.

Proof. For $x \in \Sigma_n$, let $G_x = \{y \in \Sigma_n : R(v(x), y) \leq 0\}$. The map $y \mapsto -y$ preserves π and exchanges the sets $\{y : R(v(x), y) < 0\}$ and $\{y : R(v(x), y) > 0\}$. These two sets therefore have the same Gibbs mass, and

$$\pi(G_x) = \frac{1 + \pi\{y : R(v(x), y) = 0\}}{2} \geq \frac{1}{2}.$$

Let ν_x be a probability measure on Σ_n for each x. If $\|\nu_x - \pi\|_{\mathrm{TV}} \leq \tau$, then $\nu_x(G_x) \geq \pi(G_x) - \tau \geq 1/2 - \tau$. Multiplying by $\pi(x)$ and summing over x, we obtain

$$\Big(\frac{1}{2} - \tau\Big)\, \pi\big\{x : \|\nu_x - \pi\|_{\mathrm{TV}} \leq \tau\big\} \leq \sum_x \pi(x)\, \nu_x(G_x). \tag{12.8}$$

The half-space G_x is determined by the initial state x alone. The bound (12.8) is a statement about the law of the dynamics started at a fixed state, averaged over that state afterward.

Let $(X_a)_{a \geq 0}$ be the Markov chain with kernel Q and $X_0 \sim \pi$, and set $g_k = \sum_x \pi(x) Q^k(x, G_x) = \mathbb{P}\{R(v(X_0), X_k) \leq 0\}$. Almost surely, $Q(X_{a-1}, X_a) > 0$ for every $a \geq 1$. For $k \leq m$, the hypothesis then implies that $\{R(v(X_0), X_k) \leq 0\} \subseteq \{X_a \in V \text{ for some } 0 \leq a \leq k\}$, up to a null set. Since $\pi Q = \pi$, the law of X_a is π for every deterministic a, and by a union bound we have

$$g_k \leq \sum_{a=0}^{k} \mathbb{P}\{X_a \in V\} = (k+1)\pi(V) \qquad (0 \leq k \leq m). \tag{12.9}$$

Together with (12.8) for $\nu_x = Q^k(x,\cdot)$, this proves (12.6). The matrices ntQ and ntI commute, which gives $P_t = \sum_{k \geq 0} \mathbb{P}\{\mathsf{N} = k\} Q^k$ and

$$\sum_x \pi(x) P_t(x, G_x) = \sum_{k \geq 0} \mathbb{P}\{\mathsf{N} = k\}\, g_k \leq (m+1)\pi(V) + \mathbb{P}\{\mathsf{N} > m\},$$

where we used (12.9) for $k \leq m$ and $g_k \leq 1$ for $k > m$. With (12.8) for $\nu_x = P_t(x,\cdot)$, this proves (12.7). □

The union bound (12.9) uses stationarity only at the deterministic times $0, \dots, k$, and it does not use independence along the chain. The set V and the map v are fixed before the chain is run.

We now apply Lemma 12.4 to the heat-bath kernel at a fixed disorder. Recall the time scale L and the budget T_n of (12.5).

Proposition 12.5 (Gibbs mass of initial states close to equilibrium)**.** *Assume* (11.1)*, let A satisfy* (12.4)*, and let $L = An^\lambda$ and $T_n = \lceil n^2 e^L \rceil$. There exists n_0, depending only on the fixed parameters, such that the following holds for every $n \geq n_0$. Let J be a disorder for which there exists a family of reference configurations with properties (P1)–(P4) and with $\pi^J(V) \leq 3e^{-4L}$. Then for every $\tau \in [0, 1/2)$,*

$$\pi^J\{x : d^J(x, e^L) \leq \tau\} \leq \frac{3(T_n + 1)e^{-4L} + 1/n}{1/2 - \tau}, \tag{12.10}$$

$$\pi^J\{x : d^J_{\mathrm{disc}}(x, \lfloor e^L \rfloor) \leq \tau\} \leq \frac{3(T_n + 1)e^{-4L}}{1/2 - \tau}. \tag{12.11}$$

Proof. Let n_0 be so large that the conclusions of Lemma 11.1 and Lemma 12.3 hold for $n \geq n_0$, and fix a family as in the statement. We apply Lemma 12.4 with $Q = K^J$, the Gibbs measure $\pi = \pi^J$, the map $v(x)$ of Definition 11.2, the set V of the family, and $m = T_n$. The kernel K^J preserves π^J by Lemma 3.2(a), and $\pi^J(-x) = \pi^J(x)$ because $H_n(-x) = H_n(x)$. If $K^J(x_{a-1}, x_a) > 0$, then x_a and x_{a-1} differ in at most one coordinate by (3.11). Every sequence in the hypothesis of Lemma 12.4 is a path in the sense of (11.2). For $k \leq T_n$, we have $C_n(k) < 1$ by Lemma 12.3, and Proposition 12.2, applied to the path $x_0, \dots, x_k$, implies that $R(v(x_0), x_k) > 0$ when the path avoids V. We have verified the hypothesis of Lemma 12.4.

In continuous time, we have $P^J_t = e^{nt(K^J - I)}$ by (3.12), and we take $t = e^L$. Then N has mean ne^L, and Markov's inequality gives

$$\mathbb{P}\{\mathsf{N} > T_n\} \leq \frac{ne^L}{T_n} \leq \frac{ne^L}{n^2 e^L} = \frac{1}{n}. \tag{12.12}$$

Inserting (12.12) and $\pi^J(V) \leq 3e^{-4L}$ into (12.7), we obtain (12.10). In discrete time, $k = \lfloor e^L \rfloor \leq T_n$, and by (12.6) we have

$$\pi^J\{x : d^J_{\mathrm{disc}}(x, \lfloor e^L \rfloor) \leq \tau\} \leq \frac{(\lfloor e^L \rfloor + 1)\, 3e^{-4L}}{1/2 - \tau},$$

which implies (12.11). □

For every fixed $A > 0$, the right sides of (12.10) and (12.11) tend to zero. Indeed, $T_n + 1 \leq n^2 e^L + 2$, and we have

$$3(T_n + 1)e^{-4L} \leq 3n^2 e^{-3An^\lambda} + 6e^{-4An^\lambda} \longrightarrow 0. \tag{12.13}$$

12.5 The parameters and the exponents

We first list the parameters of the proof of Theorem 1.1 in the order in which they are chosen. All of them are fixed before $n \to \infty$ and before the threshold τ.

Step	Parameter	Condition and reference
1	the parameters of Table 10.1 up to H: q_β, q_*, b_1, b_2, t_0, J_β, I_β, m_β, ϑ, E, g, $\varkappa$, C_b, R, the constants of the static estimates, and H	Section 10.2; C_b from Theorem 4.11
2	the threshold geometry: intervals of length ρ_n and spacing $3\rho_n$	(11.3)
3	the block length B	any integer $B \geq H+2$, (10.17)
4	the constants c_s, C_s of the list sizes	(10.24)
5	the coefficient A_β	$0 < A_\beta < t_0 \log 25/(3BR)$, (12.4)
6	for each n: ρ_n, ε_n, η_n, $L_n = A_\beta n^{7/30}$, M, N_{lev}, the list sizes K_s, and T_n	(10.16), (11.3), (10.24), (12.5)
7	the threshold $\tau \in (0,1/2)$	enters only through $1/(1/2-\tau)$

Table 12.1. The parameters in the proof of Theorem 1.1, in the order in which they are chosen. Each parameter of Steps 1–5 depends only on β and on the parameters chosen before it.

We use the exponents $\lambda = 7/30$ and $\xi = 5/21$. Table 12.2 lists every condition on the exponents used in the proof, with its value at these exponents; the corresponding powers of n are listed in Table 10.2.

The conditions of Table 12.2 also explain why the argument gives no exponent of $1/4$ or more. For an exponent a of the mean-bias bound and an exponent $\lambda \leq a/3$ with $\lambda < 1/4$, all rows hold if ξ is chosen strictly between λ and $\min\{1/4, 3\lambda/2\}$, which is the condition (10.10); the choice $\xi = 5/21$ works for $\lambda = 7/30$. At $\lambda = 1/4$, the third row fails, since then $n\rho_n^4 = R^4$ does not tend to infinity.

12.6 Proof of Theorem 1.1

Section 11 and Sections 12.2 to 12.4 were written for a general exponent λ. They use the deterministic assumptions stated there and the scale conditions (11.1), with $\rho_n = Rn^{-\lambda}$. Section 10 is written for exponents satisfying (10.10) and an upper bound (10.11) on the mean bias, and the static estimate Lemma 9.7 is stated for all scales satisfying the conditions (S1)–(S5). In the proof below, the mean-bias input is the bound $0 \leq P_\beta - n^{-1}\mathbb{E}F_n \leq C_b n^{-7/10}$ of Theorem 4.11; the variance bound of Theorem 4.10 enters through Section 9, and the trial bound of Proposition 4.12 through Section 8.

Condition (where it is used)	Value at $\lambda = 7/30$
$3\lambda \le a$, where Cn^{-a} bounds the mean bias (denominator margin, Section 9)	$a = 7/10 = 3\lambda$
$7/15 - 2 + 6\lambda < 0$ (Chebyshev bound for the denominator)	$-2/15$
$1 - 4\lambda > 0$ (local and separated-overlap numerators, $e^{-cn\rho_n^4}$)	$1/15$
$1 - 4\xi > 0$ (fixed-cutoff numerators, $e^{-cn\eta_n^4}$)	$1/21$
$3\lambda - 2\xi > 0$ (so that $\rho_n^3 = o(\eta_n^2)$)	$47/210$
$1 - 3\lambda > \lambda$ (conditional narrow mass $e^{-cn\rho_n^3}$ against e^{CL_n})	$3/10 > 7/30$
$1 - 2\xi > \lambda$ (conditional fixed-cutoff mass against e^{CL_n})	$11/21 > 7/30$
$\lambda < 1$ (overlap-neighborhood bound needs $L_n = o(n)$; $1/(n\rho_n) \to 0$)	yes
$\xi - \lambda > 0$ (so that $\eta_n = o(\rho_n)$, Section 11)	$1/210$

Table 12.2. The conditions on the exponents, with $\xi = 5/21$. The entries in the last column are the values of the left sides at $\lambda = 7/30$; for instance, the probability that the denominator bound fails is $O(n^{7/15-2+6\lambda}) = O(n^{-2/15})$, and the probability that the bound for a single local numerator fails is at most $e^{-cn^{1-4\lambda}} = e^{-cn^{1/15}}$. In the first row, the mean-bias bound of Theorem 4.11 has $a = 3\lambda$, and the strict margin in (S1) comes from the choice of R in (10.15).

Proof of Theorem 1.1. Fix $\beta > 1$. We take the exponents $\lambda = 7/30$ and $\xi = 5/21$ in (10.10), and the mean-bias input (10.11) is the bound of Theorem 4.11, with constant C_b. We fix the parameters of Steps 1–4 of Table 12.1 as in Sections 10.2 and 10.3, with $R^3 > 3C_b/(\varkappa g)$ as in (10.15) and with an integer $B \ge H + 2$ (for instance $B = H + 2$). Then we fix A_β as in Step 5, which is (12.4) with $A = A_\beta$ and $E' = 0$, and we set $L_n = A_\beta n^{7/30}$. The scales are $\rho_n = Rn^{-7/30}$, $\varepsilon_n = E\rho_n$, and $\eta_n = n^{-5/21}$. They satisfy (11.1) with $E' = 0$, since ρ_n, $\eta_n/\rho_n = n^{-1/210}/R$, and $1/(n\rho_n) = n^{-23/30}/R$ tend to zero.

We have $N_{\mathrm{lev}} \le t_0/(3\rho_n) = t_0 n^{7/30}/(3R) \le e^{L_n}$ for all sufficiently large n. By Remark 9.8(a), the scales satisfy (S1)–(S5), so that Lemma 9.7 and therefore Lemma 10.2 apply at these scales. Since $L_n = A_\beta n^\lambda$, the last assertion of Lemma 10.4 shows that $L_n \to \infty$, $L_n/n \to 0$, and (10.22) holds. By Lemma 10.4, applied with the time scale L_n, there exist measurable disorder events $\mathcal{G}_n$ with $\mathbb{P}_J(\mathcal{G}_n) \to 1$ such that, for all sufficiently large n and every $J \in \mathcal{G}_n$, there exists a family of reference configurations with N_{lev} levels, with properties (i)–(iv) of that lemma, and with $\pi^J(V) \le 3e^{-4L_n}$; as explained in Section 11.1, it has properties (P1)–(P4). The family may depend on J in an arbitrary way; it is not required to be a measurable function of J. Fix $\tau \in (0, 1/2)$. By Proposition 12.5, for all sufficiently large n and every $J \in \mathcal{G}_n$,

$$f_n(J) = \pi^J\{x : d^J(x, e^{L_n}) \le \tau\} \le \frac{3(T_n + 1)e^{-4L_n} + 1/n}{1/2 - \tau}.$$

We write e_n for the right side. The left side does not involve the family. By Section 3.4, f_n is a Borel function of J, and $0 \le f_n \le 1$. By (3.16),

$$\mathbb{P}_{J,X_0\sim\pi^J}\{d^J(X_0, e^{A_\beta n^{7/30}}) \le \tau\} = \mathbb{E}_J f_n(J) \le e_n + \mathbb{P}_J(\mathcal{G}_n^c).$$

Here $e_n \to 0$ by (12.13), and $\mathbb{P}_J(\mathcal{G}_n^c) \to 0$. This proves (1.2). The same argument with (12.11) in place of (12.10) proves (1.3).

The coefficient A_β was chosen in Step 5, before τ, and the same coefficient serves both clocks. The threshold τ enters only through the factor $1/(1/2-\tau)$, which affects the rate of convergence and not the coefficient. □

The proof uses the two kinds of exceptional events of Sections 9 and 10 differently, as in Remark 10.3. The disorder probability $\mathbb{P}_J(\mathcal{G}_n^c)$ is used once, additively, in the last display. It includes terms that decay only polynomially or as $e^{-cn\eta_n^4}$, and it is never multiplied by a number of references, levels, or time indices. At a fixed disorder in $\mathcal{G}_n$, the Gibbs mass $\pi^J(V)$ is multiplied by the T_n+1 time indices in (12.9), and the threshold probabilities are multiplied by the T_n+1 indices in the proof of Proposition 12.2; both products are small.

Proof of Corollary 1.2. Fix $\beta > 1$, $\kappa \in (0, 7/30)$, and $\tau \in (0, 1/2)$, and let A_β be the coefficient of Theorem 1.1. Since $\kappa < 7/30$, we have $n^\kappa \le A_\beta n^{7/30}$ for all sufficiently large n. By Lemma 3.2(d), the functions $t \mapsto d^J(x,t)$ and $k \mapsto d^J_{\mathrm{disc}}(x,k)$ are nonincreasing, and therefore, for all such n,

$$\{x : d^J(x, e^{A_\beta n^{7/30}}) > \tau\} \subseteq \{x : d^J(x, e^{n^\kappa}) > \tau\},$$

and likewise in discrete time, with $\lfloor e^{n^\kappa} \rfloor \le \lfloor e^{A_\beta n^{7/30}} \rfloor$. By (3.16), the realized-start probabilities are monotone in the same way, and Theorem 1.1 proves the corollary. □

12.7 Quantitative and almost-sure forms

The proof of Theorem 1.1 gives more than the convergence of the realized-start probabilities. In this section, we record four strengthenings. First, the disorder event can be chosen with a complement of stretched-exponentially small probability. Second, on this event there exists a set of initial states of Gibbs mass exponentially close to one from which the distance from equilibrium has a lower bound tending to $1/2$ up to the time horizon e^{L_n}. Third, the same holds for the Metropolis dynamics and for other single-site dynamics with bounded rates. Fourth, the relaxation time is at least of the order of the time horizon, up to a factor $n^{7/30}$. A consequence is the almost-sure form of the barrier along the full sequence of system sizes.

Three changes in the argument are needed. First, the lower bound on F_n in the event of Lemma 9.7 is obtained from the small-deviation inequality of Paouris and Valettas for convex functions of Gaussian vectors [90, Theorem 2.1 and Remark 2.4(3)], in place of Chebyshev's inequality. Second, the tolerance η_n is a fixed small multiple of ρ_n, so that the probability bounds for the fixed-scale constraints decay on the same scale $n\rho_n^4$ as the others. Third, the sign test of Lemma 12.4 is applied to each initial state separately, and the clock tail is bounded by a Chernoff bound instead of Markov's inequality. The relaxation-time bound uses in addition that the first reference takes few values. Theorem 1.1 remains the main result; the coefficient in Theorem 12.6 below is subject to a slightly smaller bound than in (12.4) with $E' = 0$, because of the second change.

Single-site dynamics. Fix the disorder. For $x \in \Sigma_n$ and $1 \le i \le n$, we write $x^{(i)} = x^{i,-x_i}$ for the configuration obtained by reversing the ith spin. For $C_{\mathrm{rate}} \ge 1$, we let $\mathfrak{L}^J(C_{\mathrm{rate}})$ be the set of generators of the form

$$\mathcal{L}f(x) = \sum_{i=1}^{n} c_i(x)\big(f(x^{(i)}) - f(x)\big), \qquad f\colon \Sigma_n \to \mathbb{R}, \tag{12.14}$$

with rates $c_i(x) \in [0, C_{\mathrm{rate}}]$, for which π^J is stationary, that is, $\sum_x \pi^J(x)\mathcal{L}f(x) = 0$ for every f. We call $\mathcal{L}$ reversible if $\pi^J(x)c_i(x) = \pi^J(x^{(i)})c_i(x^{(i)})$ for all x and i. By (3.10), the heat-bath generator $n(K^J - I)$ has the form (12.14) with $c_i(x) = p_i(-x_i \mid x) \le 1$, and it is reversible by Lemma 3.2(a). The Metropolis kernel is

$$K^J_{\mathrm{M}}(x, x^{(i)}) = \frac{1}{n}\min\{1, e^{\beta(H_n(x^{(i)}) - H_n(x))}\}, \qquad K^J_{\mathrm{M}}(x,x) = 1 - \sum_{i=1}^{n} K^J_{\mathrm{M}}(x, x^{(i)}),$$

and $K^J_{\mathrm{M}}(x,y) = 0$ otherwise. In one step, a uniformly chosen spin is reversed with the displayed acceptance probability. The identity $\pi^J(x)K^J_{\mathrm{M}}(x,x^{(i)}) = (nZ_n)^{-1}\min\{e^{\beta H_n(x)}, e^{\beta H_n(x^{(i)})}\}$, whose right side is symmetric in x and $x^{(i)}$, shows that K^J_{M} is reversible with respect to π^J. Its continuous-time generator $n(K^J_{\mathrm{M}} - I)$ has the form (12.14) with $c_i(x) \le 1$. Heat-bath and Metropolis dynamics belong to $\mathfrak{L}^J(1)$. For a reversible $\mathcal{L}$, the spectral gap $\mathrm{gap}(\mathcal{L})$ is the infimum of $-\langle f, \mathcal{L}f\rangle_{\pi^J}$ over f with $\pi^J(f) = 0$ and $\pi^J(f^2) = 1$, where $\langle f, g\rangle_{\pi^J} = \sum_x \pi^J(x)f(x)g(x)$, and the relaxation time is $1/\mathrm{gap}(\mathcal{L})$, with the convention $1/0 = +\infty$.

Theorem 12.6 (Quantitative and almost-sure forms)**.** *Fix $\beta > 1$. There exist $A > 0$, $c > 0$, $C < \infty$, and Borel sets $\mathcal{G}_n^*$ of disorders, such that the following statements hold with $L_n = An^{7/30}$.*

(a) *For every $n \ge 2$, we have $\mathbb{P}_J\big((\mathcal{G}_n^*)^c\big) \le Ce^{-cn^{1/15}}$.*

(b) *There exists n_0 such that, for every $n \ge n_0$ and every $J \in \mathcal{G}_n^*$, there exists a set $S \subseteq \Sigma_n$ with $\pi^J(S) \ge 1 - Ce^{-cn^{7/30}}$ such that every $x \in S$ satisfies*

$$\inf_{0 \le t \le e^{L_n}} d^J(x,t) \ge \frac{1}{2} - Ce^{-cn^{7/30}}, \qquad \min_{0 \le k \le e^{L_n}} d^J_{\mathrm{disc}}(x,k) \ge \frac{1}{2} - Ce^{-cn^{7/30}},$$

and the same two inequalities hold for the Metropolis dynamics, that is, with $(K^J_{\mathrm{M}})^k$ in place of $(K^J)^k$ and $e^{nt(K^J_{\mathrm{M}} - I)}$ in place of P^J_t.

(c) *For every $C_{\mathrm{rate}} \ge 1$ there exists n_1 such that, for every $n \ge n_1$, every $J \in \mathcal{G}_n^*$, and every $\mathcal{L} \in \mathfrak{L}^J(C_{\mathrm{rate}})$, there exists a set $S_{\mathcal{L}} \subseteq \Sigma_n$ with $\pi^J(S_{\mathcal{L}}) \ge 1 - Ce^{-cn^{7/30}}$ such that*

$$\| e^{t\mathcal{L}}(x,\cdot) - \pi^J \|_{\mathrm{TV}} \ge \frac{1}{2} - Ce^{-cn^{7/30}} \qquad \text{for every } x \in S_{\mathcal{L}} \text{ and } 0 \le t \le e^{L_n}.$$

(d) *For every $C_{\mathrm{rate}} \ge 1$, the threshold $n_1 = n_1(\beta, C_{\mathrm{rate}})$ in part (c) can be chosen so that, for every $n \ge n_1$, every $J \in \mathcal{G}_n^*$, and every reversible $\mathcal{L} \in \mathfrak{L}^J(C_{\mathrm{rate}})$, the relaxation time satisfies*

$$\frac{1}{\mathrm{gap}(\mathcal{L})} \ge \frac{e^{An^{7/30}}}{Cn^{7/30}}.$$

In particular, this holds for the heat-bath and Metropolis generators $n(K^J - I)$ and $n(K^J_{\mathrm{M}} - I)$.

(e) *Let $(J^{(n)})_{n \ge 2}$ be random disorders defined on one probability space, where $J^{(n)}$ has the law of the disorder of size n and the dependence between different sizes is arbitrary. Then, almost surely, $J^{(n)} \in \mathcal{G}_n^*$ for all sufficiently large n. In particular, for every fixed $\tau \in (0, 1/2)$, almost surely,*

$$\pi^{J^{(n)}}\{x : d^{J^{(n)}}(x, e^{An^{7/30}}) > \tau\} \longrightarrow 1, \qquad \pi^{J^{(n)}}\{x : d^{J^{(n)}}_{\mathrm{disc}}(x, \lfloor e^{An^{7/30}} \rfloor) > \tau\} \longrightarrow 1,$$

and the same holds for the Metropolis dynamics.

The constants A, c, C, and n_0 depend only on β; the threshold n_1 depends in addition on C_{rate}.

Part (b) contains the conclusions of Theorem 1.1, with the coefficient A of Theorem 12.6 in place of A_β. It gives in addition the size of the exceptional sets, the uniformity over all earlier times, and a lower bound tending to $1/2$ on the distance. The threshold $1/2$ itself is not reached: the argument uses only that the half-space $\{y : R(v,y) \le 0\}$ has Gibbs mass at least $1/2$. In part (e), no independence between different system sizes is needed, because the first Borel–Cantelli lemma uses only the marginal laws.

The following table lists how the earlier statements enter the proof.

Lemma 12.7 (A disorder event with stretched-exponential failure)**.** *Fix $\beta > 1$ and let $\lambda = 7/30$. Fix the parameters of Steps 1–4 of Table 12.1, with C_b the constant of Theorem 4.11, and set*

$$\rho_n = Rn^{-\lambda}, \qquad \varepsilon_n = E\rho_n, \qquad \eta_n = \frac{\rho_n}{1000}, \qquad L_n = An^{\lambda}$$

Statement	Use in the proof of Theorem 12.6
Section 11; Lemma 12.1; Proposition 12.2	used as stated; they assume (11.1), which holds with $E' = 1/1000$
Lemma 12.3	used as stated, with $E' = 1/1000$, so that (12.4) reads $A < t_0 \log(250/11)/(3BR)$; it covers every $m \le n^3 e^{L_n}$
Lemma 9.7	statement used as stated, at $\eta_n = \rho_n/1000$; its hypotheses are verified in Lemma 12.7
Step 1 of the proof of Lemma 9.7 (Chebyshev)	replaced by the Paouris–Valettas inequality in Lemma 12.7
Lemma 10.1	used as stated, including its bound on $\mathbb{P}(\mathcal{C}_n(D_1)^c)$
Lemma 10.2	used as stated, at $\eta_n = \rho_n/1000$, in Step 4 of the proof of Lemma 12.7
Lemma 10.4	used as stated, with q_n supplied by that Step 4; (10.22) is verified there
Lemma 12.4, Proposition 12.5, (12.12)	replaced by Lemma 12.8 and the Chernoff bound (12.17)
new	Lemma 12.9 (relaxation time)

Table 12.3. The earlier statements in the proof of Theorem 12.6.

for a fixed $A > 0$. Let $\mathcal{A}_n$ be the event of Lemma 9.7 at these scales, with the data of Section 10, and let $\mathcal{G}_n^* = \mathcal{A}_n \cap \bigcap_{s<B} \mathcal{C}_n(C_s + 5)$ be the event (10.26) *at these scales. Then the following hold.*

(i) *There exist $C, c > 0$ such that $\mathbb{P}_J((\mathcal{G}_n^*)^c) \le Ce^{-cn^{1-4\lambda}}$ for all sufficiently large n.*

(ii) *For all sufficiently large n and every $J \in \mathcal{G}_n^*$, there exists a family of reference configurations with N_{lev} levels, with properties (P1)–(P4), and with $\pi^J(V) \le 3e^{-4L_n}$.*

Proof. Since $\lambda > 0$, we have $\rho_n = Rn^{-\lambda} \to 0$. We also verify the conditions (S1)–(S5) of Lemma 9.7. For (S1), we have $\Delta_{n,\beta}/\rho_n^3 \le C_b/R^3 < \varkappa g/3$, by Theorem 4.11 and (10.15). Next, $n^{7/15}/(n^2\rho_n^6) = R^{-6}n^{7/15-2+6\lambda} = R^{-6}n^{-2/15}$, which is (S2). We have $n\rho_n^4 = R^4 n^{1-4\lambda} = R^4 n^{1/15}$ and $n\eta_n^4 = 10^{-12}R^4 n^{1/15}$, which gives (S3) and (S4). Finally, $\rho_n^3/\eta_n^2 = 10^6\rho_n \to 0$, which is (S5). Therefore Lemma 9.7 applies at these scales.

Step 1: the numerators. Steps 2–4 of the proof of Lemma 9.7 bound the probability that one of the events (A2)–(A4) fails by

$$C(n+1)^3\big(e^{-cn\rho_n^4} + e^{-cn\eta_n^4}\big) + C(n+1)e^{-cn} \le C'e^{-c'n^{1-4\lambda}},$$

for all large n, where we used $n\eta_n^4 = 10^{-12}n\rho_n^4$ and $n\rho_n^4 = R^4 n^{1-4\lambda}$, and absorbed the polynomial factors by decreasing the constant in the exponent.

Step 2: the denominator. The function $F_n = \log Z_n$ of the standard Gaussian vector $J = (J_{ij})_{i<j}$ is convex. Indeed, its Hessian is $\beta^2 n^{-1}$ times the covariance matrix of the

vector $(x_i x_j)_{i<j}$ under π^J, which is positive semidefinite. It is Lipschitz by Remark 4.22, so it is square integrable. By Theorem 4.10, $\sigma_n^2 = \operatorname{Var} F_n \le C_v n^{7/15}$ for large n, with C_v the variance constant there. We write $a_0 < c_0$ for the static denominator margins in (9.19). For all sufficiently large n, $\Delta_{n,\beta} \le a_0 \rho_n^3$, so failure of (A1) implies $F_n - \mathbb{E}F_n < -u_n$, where $u_n = (c_0 - a_0) n \rho_n^3$. If $\sigma_n = 0$, this event is empty. Otherwise, let $t_n = u_n/\sigma_n$. Then

$$t_n \ge \frac{(c_0 - a_0) R^3}{C_v^{1/2}} \, n^{23/30 - 3\lambda},$$

and the exponent $23/30 - 3\lambda = 1/15$ is positive, so that $t_n \ge 2$ for large n. By [90, Theorem 2.1 and Remark 2.4(3), (2.4)], a convex square-integrable function f of a standard Gaussian vector Z satisfies

$$\mathbb{P}\Big(f(Z) - \mathbb{E}f(Z) < -t \sqrt{\operatorname{Var} f(Z)} \Big) \le \frac{1}{2} \exp\Big(-\frac{\pi}{1024} (t-1)^2 \Big) \qquad (t > 1). \tag{12.15}$$

Applying (12.15) with $t = t_n$ and using $(t_n - 1)^2 \ge t_n^2/4$, we obtain

$$\mathbb{P}\big(F_n < n(P_\beta - c_0 \rho_n^3) \big) \le \frac{1}{2} \exp\Big(-\frac{\pi t_n^2}{4096} \Big) \le \frac{1}{2} \exp\Big(-\frac{\pi (c_0 - a_0)^2 R^6}{4096 \, C_v} \, n^{23/15 - 6\lambda} \Big).$$

Since $23/15 - 6\lambda = 2/15$ is larger than $1 - 4\lambda = 1/15$, this is at most $e^{-c n^{1-4\lambda}}$ for large n. Only the full partition function enters here. For the constrained numerators, the upper tail is needed, and the bounds of Step 1 remain those of Lemma 9.7.

Step 3: the overlap-neighborhood events. Fix $D_1 > 0$. By Lemma 10.1, there exists $D_4 > 0$, depending only on β and D_1, such that, for all sufficiently large n,

$$\mathbb{P}\big(\mathcal{C}_n(D_1)^c \big) \le 3 e^{-D_4 L_n} = 3 e^{-D_4 A n^\lambda}.$$

Since $\lambda = 7/30$ is larger than $1 - 4\lambda = 1/15$, this is at most $e^{-c n^{1-4\lambda}}$ for large n. There are B such events in the definition of $\mathcal{G}_n^*$, with $D_1 = C_s + 5$ for $0 \le s < B$. Combining Steps 1–3, we obtain part (i).

Step 4: the families. The present sequences satisfy $\rho_n \to 0$ and (S1)–(S5), as we checked at the start of the proof. By Lemma 10.2, for all large n and every $J \in \mathcal{A}_n$, we have $(\pi^J)^{\otimes 3}(\mathcal{B}_n) \le q_n$ with $q_n = C(n+1)^3 (e^{-cn\rho_n^3} + e^{-cn\eta_n^2} + e^{-cn})$, where $\mathcal{B}_n$ is defined by (10.20) with the present η_n. Since $n\rho_n^3 = R^3 n^{1-3\lambda}$ and $n\eta_n^2 = 10^{-6} R^2 n^{1-2\lambda}$, and $1 - 3\lambda = 3/10$ and $1 - 2\lambda = 8/15$ are larger than $\lambda = 7/30$, we have $e^{CL_n} q_n \to 0$ for every fixed C, which is (10.22). Moreover $L_n \to \infty$, $L_n/n \to 0$, and $N_{\mathrm{lev}} \le t_0 n^\lambda / (3R) \le e^{L_n}$ for large n. Lemma 10.4 therefore provides, for all large n and every $J \in \mathcal{G}_n^*$, a family with properties (i)–(iv) of that lemma and $\pi^J(V) \le 3 e^{-4L_n}$. As in Section 11.1, these imply (P1)–(P4), with $\zeta = 2\eta_n$. This proves part (ii). □

The next lemma refines Lemma 12.4: it bounds the distance from equilibrium for each initial state separately, in terms of the probability that the chain started there visits V.

Lemma 12.8 (The sign test for each initial state)**.** *Let Q, π, v, V, and m satisfy the hypotheses of Lemma 12.4. Let $(X_a)_{a\ge 0}$ be the Markov chain with kernel Q, write $\mathbb{P}_x$ for its law when $X_0 = x$, and set*

$$\phi(x) = \mathbb{P}_x\{X_a \in V \text{ for some } 0 \le a \le m\}, \qquad G_x = \{y : R(v(x), y) \le 0\}.$$

Then the following hold.

(a) $\sum_x \pi(x)\phi(x) \le (m+1)\pi(V)$, *and hence* $\pi\{x : \phi(x) > u\} \le (m+1)\pi(V)/u$ *for every* $u > 0$.

(b) *For every x and every $0 \le k \le m$, we have $Q^k(x, G_x) \le \phi(x)$ and $\|Q^k(x,\cdot) - \pi\|_{\mathrm{TV}} \ge 1/2 - \phi(x)$.*

(c) *For every x, every $q_{\mathrm{u}} > 0$, and every $t \ge 0$, let N have the Poisson law with mean $q_{\mathrm{u}}t$. Then*

$$\|e^{q_{\mathrm{u}}t(Q-I)}(x,\cdot) - \pi\|_{\mathrm{TV}} \ge \frac{1}{2} - \phi(x) - \mathbb{P}\{\mathsf{N} > m\},$$
$$\sum_x \pi(x) e^{q_{\mathrm{u}}t(Q-I)}(x, G_x) \le (m+1)\pi(V) + \mathbb{P}\{\mathsf{N} > m\}.$$

Proof. For part (a), we have $\sum_x \pi(x)\phi(x) = \mathbb{P}_\pi\{X_a \in V \text{ for some } a \le m\}$, where $\mathbb{P}_\pi$ is the law of the chain with $X_0 \sim \pi$. Since $\pi Q = \pi$, the law of X_a under $\mathbb{P}_\pi$ is π for every deterministic a, and a union bound over $0 \le a \le m$ gives the first assertion, as in (12.9). The second assertion is Markov's inequality. For part (b), almost surely under $\mathbb{P}_x$ we have $Q(X_{a-1}, X_a) > 0$ for every $a \ge 1$. By the hypothesis of Lemma 12.4, on the event that $X_0, \dots, X_k$ avoid V we have $R(v(x), X_k) > 0$, that is, $X_k \notin G_x$. It follows that $Q^k(x, G_x) \le \mathbb{P}_x\{X_a \in V \text{ for some } a \le k\} \le \phi(x)$. As in the proof of Lemma 12.4, spin-flip symmetry gives $\pi(G_x) \ge 1/2$, and therefore $\|Q^k(x,\cdot) - \pi\|_{\mathrm{TV}} \ge \pi(G_x) - Q^k(x, G_x) \ge 1/2 - \phi(x)$. For part (c), the matrices $q_{\mathrm{u}}tQ$ and $q_{\mathrm{u}}tI$ commute, so $e^{q_{\mathrm{u}}t(Q-I)} = \sum_{k\ge 0} \mathbb{P}\{\mathsf{N} = k\}Q^k$. By part (b),

$$e^{q_{\mathrm{u}}t(Q-I)}(x, G_x) \le \sum_{k \le m} \mathbb{P}\{\mathsf{N} = k\}\phi(x) + \mathbb{P}\{\mathsf{N} > m\} \le \phi(x) + \mathbb{P}\{\mathsf{N} > m\},$$

which gives the first inequality as in part (b). Multiplying by $\pi(x)$, summing over x, and using part (a), we obtain the second. □

The next lemma combines the standard continuous-time L^2 spectral-gap contraction [68, Lemma 20.5] with a partition of the initial states according to their chosen reference. This grouping gives a bound in terms of the number of references.

Lemma 12.9 (Relaxation time and a sign test with few references)**.** *Let $\mathcal{L}$ be a generator on Σ_n that is reversible with respect to a probability measure π with $\pi(-x) = \pi(x)$, and let $\operatorname{gap}(\mathcal{L})$ be its spectral gap. Let $v\colon \Sigma_n \to \Sigma_n$ take at most N_v values, and let $G_x = \{y : R(v(x), y) \le 0\}$. If $t > 0$ and*

$$\omega = \sum_x \pi(x) e^{t\mathcal{L}}(x, G_x) < \frac{1}{2},$$

then $\operatorname{gap}(\mathcal{L})\, t \le (\log N_v)/2 - \log(1 - 2\omega)$.

Proof. For f with $\pi(f) = 0$, the function $f_t = e^{t\mathcal{L}} f$ also has mean zero, since π is stationary, and the derivative of $\langle f_t, f_t\rangle_\pi$ in t is $2\langle f_t, \mathcal{L} f_t\rangle_\pi \le -2\operatorname{gap}(\mathcal{L})\langle f_t, f_t\rangle_\pi$, by the definition of the spectral gap. Gronwall's inequality gives $\|e^{t\mathcal{L}} f\|_{L^2(\pi)} \le e^{-\operatorname{gap}(\mathcal{L})t}\|f\|_{L^2(\pi)}$.

For each value $\widetilde{v}$ of v, let $U_{\widetilde{v}} = \{x : v(x) = \widetilde{v}\}$ and $H_{\widetilde{v}} = \{y : R(\widetilde{v}, y) \le 0\}$. The sets $U_{\widetilde{v}}$ partition Σ_n, and $G_x = H_{\widetilde{v}}$ for $x \in U_{\widetilde{v}}$. As in the proof of Lemma 12.4, spin-flip symmetry gives $\pi(H_{\widetilde{v}}) \ge 1/2$, and therefore

$$\sum_{\widetilde{v}} \pi(U_{\widetilde{v}})\pi(H_{\widetilde{v}}) \ge \frac{1}{2}, \qquad \sum_{\widetilde{v}} \langle \mathbf{1}_{U_{\widetilde{v}}}, e^{t\mathcal{L}} \mathbf{1}_{H_{\widetilde{v}}}\rangle_\pi = \sum_x \pi(x) e^{t\mathcal{L}}(x, G_x) = \omega.$$

Since $e^{t\mathcal{L}}\mathbf{1} = \mathbf{1}$ and π is stationary, $\langle \mathbf{1}_U, e^{t\mathcal{L}}\mathbf{1}_{U'}\rangle_\pi - \pi(U)\pi(U') = \langle \mathbf{1}_U - \pi(U), e^{t\mathcal{L}}(\mathbf{1}_{U'} - \pi(U'))\rangle_\pi$ for all sets $U, U' \subseteq \Sigma_n$. Using the Cauchy–Schwarz inequality, the contraction just proved, $\|\mathbf{1}_U - \pi(U)\|^2_{L^2(\pi)} = \pi(U)(1 - \pi(U)) \le \pi(U)$, and $\|\mathbf{1}_{U'} - \pi(U')\|_{L^2(\pi)} \le 1/2$ for $U = U_{\widetilde{v}}$ and $U' = H_{\widetilde{v}}$, we find

$$\frac{1}{2} - \omega \le \sum_{\widetilde{v}} \bigl(\pi(U_{\widetilde{v}})\pi(H_{\widetilde{v}}) - \langle \mathbf{1}_{U_{\widetilde{v}}}, e^{t\mathcal{L}}\mathbf{1}_{H_{\widetilde{v}}}\rangle_\pi\bigr) \le \frac{e^{-\operatorname{gap}(\mathcal{L})t}}{2} \sum_{\widetilde{v}} \sqrt{\pi(U_{\widetilde{v}})} \le \frac{e^{-\operatorname{gap}(\mathcal{L})t}}{2}\sqrt{N_v}.$$

In the last step, we used the Cauchy–Schwarz inequality and $\sum_{\widetilde{v}} \pi(U_{\widetilde{v}}) = 1$. Rearranging gives the claim. □

The bound depends on the number N_v of values of v, not on the size of Σ_n. If v took all 2^n values, the lemma would only give $\operatorname{gap}(\mathcal{L})\, t \le Cn$. The first reference takes at most e^{CL_n} values, and this gives the factor $n^{7/30}$, rather than n, in Theorem 12.6(d).

Proof of Theorem 12.6. Fix $\beta > 1$ and let $\lambda = 7/30$. We fix the parameters of Steps 1–4 of Table 12.1 as in Sections 10.2 and 10.3, with C_b the constant of Theorem 4.11, with $R^3 > 3C_b/(\varkappa g)$, and with an integer $B \ge H + 2$. We set $E' = 1/1000 = E/10$, $\eta_n = \rho_n/1000$, and fix A with

$$0 < A < \frac{t_0 \log(250/11)}{3BR},$$

which is (12.4) with $E' = 1/1000$, since $4(E + E') = 11/250$. Let $L_n = An^\lambda$, let $\mathcal{G}_n^*$ be the event of Lemma 12.7, and let $m_* = \lfloor n^3 e^{L_n} \rfloor$, a budget of update attempts. Throughout, n is large enough that the conclusions of Lemmas 11.1, 12.3, and 12.7 hold; this fixes n_0.

Part (a). By Lemma 12.7(i) and $1 - 4\lambda = 1/15$, the bound holds for $n \geq n_0$, and we enlarge C so that $Ce^{-cn^{1/15}} \geq 1$ for $2 \leq n < n_0$.

The obstruction. Fix $n \geq n_0$ and $J \in \mathcal{G}_n^*$, and fix a family as in Lemma 12.7(ii). The scales satisfy (11.1) with $E' = 1/1000$. By Lemma 11.1, the margins (11.6) hold, and by Lemma 12.3, $C_n(m) < 1$ for every $m \leq m_*$. By Proposition 12.2, every path $x_0, \dots, x_m$ with $m \leq m_*$ that avoids V satisfies $R(v(x_0), x_m) > 0$.

Parts (b) and (c). We use the uniformization of a bounded-rate generator into a discrete kernel and a Poisson clock [68, Section 20.1]. Let $\mathcal{L} \in \mathfrak{L}^J(C_{\mathrm{rate}})$ and set $Q_{\mathcal{L}} = I + \mathcal{L}/(C_{\mathrm{rate}}n)$. Its off-diagonal entries are $Q_{\mathcal{L}}(x, x^{(i)}) = c_i(x)/(C_{\mathrm{rate}}n) \geq 0$, its diagonal entries are $1 - \sum_i c_i(x)/(C_{\mathrm{rate}}n) \geq 0$, and $Q_{\mathcal{L}}(x, y) = 0$ unless $y = x$ or $y = x^{(i)}$ for some i. We see that $Q_{\mathcal{L}}$ is a Markov kernel, π^J is stationary for it, every sequence with $Q_{\mathcal{L}}(x_{a-1}, x_a) > 0$ is a path in the sense of (11.2), and $e^{t\mathcal{L}} = e^{C_{\mathrm{rate}}nt(Q_{\mathcal{L}} - I)}$. For the heat-bath and Metropolis generators with $C_{\mathrm{rate}} = 1$, we have $Q_{\mathcal{L}} = K^J$ and $Q_{\mathcal{L}} = K_{\mathrm{M}}^J$, respectively. Since $\pi^J(-x) = \pi^J(x)$, the obstruction shows that the hypotheses of Lemma 12.4 hold for $Q = Q_{\mathcal{L}}$, $\pi = \pi^J$, the map v, the set V of the family, and $m = m_*$. Let $\phi_{\mathcal{L}}$ be the function ϕ of Lemma 12.8 for this kernel, and let

$$S_{\mathcal{L}} = \{x : \phi_{\mathcal{L}}(x) \leq e^{-L_n}\}.$$

Since $m_* + 1 \leq 2n^3 e^{L_n}$ and $\pi^J(V) \leq 3e^{-4L_n}$, by Lemma 12.8(a), we have

$$\pi^J(S_{\mathcal{L}}^c) \leq e^{L_n}(m_* + 1)\pi^J(V) \leq 6n^3 e^{-2L_n}. \tag{12.16}$$

For $x \in S_{\mathcal{L}}$ and $0 \leq k \leq e^{L_n} \leq m_*$, by Lemma 12.8(b), we have $\|Q_{\mathcal{L}}^k(x, \cdot) - \pi^J\|_{\mathrm{TV}} \geq 1/2 - e^{-L_n}$. For continuous time, the bound at $t = 0$ is immediate. Let $0 < t \leq e^{L_n}$ and let N have the Poisson law with mean $\mu_{\mathrm{P}} = C_{\mathrm{rate}}nt \leq C_{\mathrm{rate}}ne^{L_n}$. For every integer $j \geq \mu_{\mathrm{P}}$, Markov's inequality applied to z^{N} with $z = j/\mu_{\mathrm{P}} \geq 1$, together with $\mathbb{E}z^{\mathsf{N}} = e^{\mu_{\mathrm{P}}(z-1)}$, gives $\mathbb{P}\{\mathsf{N} \geq j\} \leq z^{-j}e^{\mu_{\mathrm{P}}(z-1)} = e^{-\mu_{\mathrm{P}}}(e\mu_{\mathrm{P}}/j)^j$. With $j = m_* + 1 \geq n^3 e^{L_n}$, this implies

$$\mathbb{P}\{\mathsf{N} > m_*\} \leq \Big(\frac{e\, C_{\mathrm{rate}}ne^{L_n}}{n^3 e^{L_n}}\Big)^{m_*+1} = \Big(\frac{eC_{\mathrm{rate}}}{n^2}\Big)^{m_*+1} \leq e^{-(m_*+1)} \leq e^{-L_n} \tag{12.17}$$

once $n^2 \geq e^2 C_{\mathrm{rate}}$. By Lemma 12.8(c) with $q_{\mathrm{u}} = C_{\mathrm{rate}}n$, every $x \in S_{\mathcal{L}}$ satisfies $\|e^{t\mathcal{L}}(x, \cdot) - \pi^J\|_{\mathrm{TV}} \geq 1/2 - 2e^{-L_n}$ for $0 \leq t \leq e^{L_n}$. Taking $n_1 \geq n_0$ so large that also $n_1^2 \geq e^2 C_{\mathrm{rate}}$ and $12n^3 e^{-2L_n} \leq e^{-L_n}$ for $n \geq n_1$, we obtain part (c) with $c = A$ and $C = 2$. For part (b), we apply this with $C_{\mathrm{rate}} = 1$ to the heat-bath and Metropolis generators, for which $Q_{\mathcal{L}}$ is K^J and K_{M}^J, and take $S = S_{n(K^J - I)} \cap S_{n(K_{\mathrm{M}}^J - I)}$. By (12.16), $\pi^J(S^c) \leq 12n^3 e^{-2L_n} \leq e^{-L_n}$, and the four distance bounds hold on S, with $\lfloor e^{L_n} \rfloor$ discrete steps and $1/2 - 2e^{-L_n}$ as the lower bound. Enlarging n_0 to the value of n_1 for $C_{\mathrm{rate}} = 1$ completes part (b).

Part (d). Let $\mathcal{L} \in \mathfrak{L}^J(C_{\mathrm{rate}})$ be reversible, and let $t = e^{L_n}$. By Lemma 12.8(c) with $q_{\mathrm{u}} = C_{\mathrm{rate}} n$, by (12.17), and by $\pi^J(V) \le 3e^{-4L_n}$,

$$\omega = \sum_x \pi^J(x) e^{t\mathcal{L}}(x, G_x) \le 6n^3 e^{-3L_n} + e^{-L_n} \le \frac{1}{4}$$

for $n \ge n_1$, after enlarging the threshold n_1 of part (c) by an amount depending only on β and C_{rate}. Write $\widehat{c} = c(5) > 0$ for the first list-size constant in (10.24). The map v takes values among the signed entries of the list $\mathcal{F}_0$, which has $K_0 = \lceil e^{(\widehat{c}+1)L_n} \rceil$ entries by (10.24); hence it takes at most $N_v = 2K_0 \le 4e^{(\widehat{c}+1)L_n}$ values, where $\widehat{c}$ depends only on β. By Lemma 12.9,

$$\operatorname{gap}(\mathcal{L}) e^{L_n} \le \frac{1}{2} \log\bigl(4e^{(\widehat{c}+1)L_n}\bigr) + \log 2 \le (\widehat{c}+2) L_n$$

for $n \ge n_1$, after enlarging n_1 once more by an amount depending only on β. If $\operatorname{gap}(\mathcal{L}) = 0$, the conclusion of part (d) is immediate by the convention $1/0 = +\infty$. Otherwise we may divide by the positive gap, and we obtain $1/\operatorname{gap}(\mathcal{L}) \ge e^{L_n}/((\widehat{c}+2)L_n) = e^{An^\lambda}/((\widehat{c}+2)An^\lambda)$, which is part (d) after enlarging C. The enlarged n_1 also serves in part (c).

Part (e). By part (a), $\sum_n \mathbb{P}(J^{(n)} \notin \mathcal{G}_n^*) \le \sum_n Ce^{-cn^{1/15}} < \infty$. By the first Borel–Cantelli lemma, which uses only the marginal laws of the $J^{(n)}$, almost surely $J^{(n)} \in \mathcal{G}_n^*$ for all large n. On this event, part (b) gives, for all large n, a set of Gibbs mass at least $1 - Ce^{-cn^\lambda}$ on which $d^{J^{(n)}}(x, e^{An^\lambda})$ and $d^{J^{(n)}}_{\mathrm{disc}}(x, \lfloor e^{An^\lambda} \rfloor)$ are at least $1/2 - Ce^{-cn^\lambda}$, which exceeds τ for large n, and likewise for the Metropolis dynamics. This completes the proof. □

PART IV

The upper bound

SECTION 13

Primitive disorder, the scalar process beyond the support, and the parameters

Throughout Part IV, we fix $\beta > 1$. This section sets the notation for that Part. We identify the primitive Gaussian disorder with the completed model of Section 3, record the Gaussian matrix estimates used in Sections 14, 18, and 19, and give the exponent parameters. We then continue the scalar diffusion of Section 4 beyond the support of the Parisi measure. The scalar profiles and the one-row law defined below are used throughout the construction of the proposal and its analysis. We follow the scalar construction of [85, Sections 2–3].

The notation in Table 13.1 is local to this Part. All other notation retains the meaning fixed in Sections 3 and 4. In particular, $\mathbb{P}$ denotes probability, P_β is the limiting pressure, and all exponents are written in sans serif.

13.1 Primitive disorder and the completed Hamiltonian

Let γ_n be standard Gaussian measure on $\mathbb{R}^{n\times n}$. For $g \in \mathbb{R}^{n\times n}$ and $x \in \Sigma_n$, define

$$\begin{aligned} W(g) &= \frac{g+g^{\mathsf{T}}}{\sqrt{2n}}, \qquad H_g^c(x) = \frac{1}{2}x^{\mathsf{T}}W(g)x, \\ F^c(g) &= \log \sum_{x\in\Sigma_n} e^{\beta H_g^c(x)}, \qquad \pi_g(x) = e^{\beta H_g^c(x) - F^c(g)}. \end{aligned} \tag{13.1}$$

The dictionary to Section 3.2 is

$$J_{ij}(g) = \frac{g_{ij}+g_{ji}}{\sqrt{2}} \quad (i<j), \qquad J'_{ii}(g) = g_{ii}, \qquad A_{ij}(g) = \frac{g_{ij}-g_{ji}}{2} \quad (i<j).$$

We note that the variables J_{ij} and J'_{ii} are independent standard Gaussians. The antisymmetric coordinates A_{ij} are independent $N(0,1/2)$ variables, independent of $W(g)$. Expanding the quadratic form, we find

$$H_g^c(x) = H_n^{J(g)}(x) + G(g), \qquad G(g) = \frac{1}{2}\operatorname{tr} W(g) = \frac{1}{\sqrt{2n}}\sum_i g_{ii} \sim N(0,1/2).$$

Thus $F^c(g) = F_n^c$, and Lemma 3.1(b) implies that $\pi_g = \pi^{J(g)}$. We write $\mathcal{G}_n = W_*\gamma_n$ for the GOE law with off-diagonal variance $1/n$ and diagonal variance $2/n$, and $\mathrm{Unif}_n(x) = 2^{-n}$ for the uniform law on Σ_n.

Table 13.1. Notation local to Part IV. The right column records the meanings used here and any earlier meanings that they replace.

Symbols	Meaning in this Part
U, m	Scalar potential and its first derivative in variance time; these replace $\Phi_{\alpha_\beta}, u_{\alpha_\beta}$ after the time change.
Y, M	Scalar diffusion and its Wiener driver in variance time; Y replaces the overlap-time diffusion of Section 4.
B, D	Reflected Wiener pair; B replaces the driver denoted B in overlap time.
δ, ϵ_n	Smoothing variance $n^{-\mathsf{D}}$ and cut-balance threshold $n^{-\mathsf{d}}$.
h, h_j	Mesh lengths; fields involving the completed matrix are written $(W(g)x)_i$.
ρ, σ	Deterministic reflection coefficients, replacing the overlap scale ρ and the resampled spin symbol σ.
θ	Rotation angle, or an exponential-moment parameter in the separate concentration arguments.
p, r, e, j	Scalar profiles, with e_h, b_j, s_j, w_j their mesh quantities; the subscripted probabilities p_i^J keep their meaning.
R_V, R_h, R_n, R_D	Inverse-profile bound, mesh residual, enlargement radius, and reference law, respectively; $R(x, y)$ remains the overlap.
H, H_0, E_H	Query cross-moment matrix, its population value, and its error; Hamiltonians carry their superscripts or subscripts.
H_*	Maximum absolute spin energy in Section 15, and a Bernstein scale in Section 16.
$\mu_0^{\mathrm{src}}, \mu_0^{\mathrm{rep}}$	Unlinked source and replay laws.
$\Lambda^{\mathrm{src}}, \Lambda^{\mathrm{rep}}$	Linked source and replay laws.
Q_V, Q_F, ν_V	Output law, fine output law, and one-row law.
L	Spin likelihood in Sections 14 and 15; $L = \mathrm{Up}_{1/2}(E_A)$ in Sections 16 and 17.
$W_{\mathrm{wt}}, W_0, W_2$	Weighted error size, ambient GOE matrix, and 2-Wasserstein distance; $W(g)$ is the completed matrix.
$\mathcal{C}_n, \mathcal{H}_n, \mathcal{W}_n, \mathrm{Bad}$	Events and spin sets defined where they are used; $\mathcal{H}_n$ denotes the pressure event.

For the dynamics, let $P_t^g = P_t^{J(g)}$. The rate of a spin flip is

$$p_i^{J(g)}(-x_i \mid x) = \frac{\pi_g(x^{i,-x_i})}{\pi_g(x) + \pi_g(x^{i,-x_i})}. \tag{13.2}$$

The generator is $\mathcal{L}^{J(g)}$, with rate one per site as in (3.12). For a cut $S \subseteq \Sigma_n$, its Dirichlet flow is

$$\mathcal{Q}_g(S, S^c) = \sum_{x \in S} \sum_{i=1}^{n} \pi_g(x) p_i^{J(g)}(-x_i \mid x) \mathbf{1}\{x^{i,-x_i} \notin S\}. \tag{13.3}$$

The diagonal of $W(g)$ adds the same constant to all spin energies, so it cancels from the Gibbs measure and the dynamics. We compare proposals with the joint law

$$\Pi_n(\mathrm{d}g, x) = \gamma_n(\mathrm{d}g)\pi_g(x). \tag{13.4}$$

Its $(J(g), x)$-marginal is $\mathbb{P}_{J, X_0 \sim \pi^J}$ from Definition 3.3. In the transition law, the spin sampled from π_g is held fixed.

The relative entropy is $\mathrm{KL}(\mu \| \nu) = \int \log(\mathrm{d}\mu / \mathrm{d}\nu)\, \mathrm{d}\mu$ when $\mu \ll \nu$, where the integral takes values in the extended nonnegative reals, and $+\infty$ otherwise. In finite sums, $0 \log 0 = 0$, and a positive numerator over zero contributes $+\infty$. We use the asymptotic conventions of Section 3.5. In every eventual bound, we first fix $\beta > 1$ and all exponents, then choose finite constants and n_0, and then assert the bound for every integer $n \geq n_0$. Constants $C_\beta, c_\beta > 0$ may change between inequalities. They depend only on β and, when an exponent is absorbed, on the displayed fixed exponents; they are independent of n, the selected cut, and the rotation angle. The bounds are for each fixed temperature and saving $\omega < 1/4$.

Conventions. Spaces of continuous paths carry the Borel sigma algebra of the uniform topology, products carry the product sigma algebras, and Wiener measures are supported on paths starting at zero. A bridge of duration $h > 0$ has the law Br_h obtained from Wiener measure by $b(t) \mapsto b(t) - (t/h)b(h)$. These spaces are standard Borel. We complete filtrations when constructing stochastic integrals and then choose Borel versions on the original spaces. All refinements defining integrals are taken at fixed n, fixed horizon, and fixed query mesh, before $n \to \infty$. Versions are identified outside fixed-n null sets. For the terminal spin likelihood L defined in (14.6), the bound $L \leq 2^n$ is used only to transfer fixed-n null sets and integral versions. All probability estimates are proved under the law to which they apply.

13.2 Gaussian orthogonal ensemble facts

We record the Gaussian facts needed for the adaptive construction of the matrix in Lemma 14.1 and for the exceptional events of the disorder in Sections 18 and 19. The

successive conditioning on orthogonal queries is the Gaussian-revelation method used in [25, Sections 2–3] and in the proposal of [85, Section 4.1].

Lemma 13.1 (Gaussian conditioning)**.** *Let $W \sim \mathcal{G}_n$. Its law is invariant under orthogonal conjugation. In an orthonormal basis, the entries on and above the diagonal are independent, with variances $2/n$ on the diagonal and $1/n$ off the diagonal. After revealing its columns in a fixed subspace, the compression to the orthogonal complement is an independent GOE block with the original variance normalization $1/n$. The same conditional assertion holds for successive orthogonal unit query directions chosen measurably from the previously revealed columns and independent initial data.*

Proof. For deterministic vectors u, v, u', v', the covariance is

$$\mathbb{E}[(u^{\mathsf{T}} W v)(u'^{\mathsf{T}} W v')] = \frac{1}{n}\big[(u \cdot u')(v \cdot v') + (u \cdot v')(v \cdot u')\big].$$

This follows by expanding (13.1). Since W is Gaussian, this covariance implies that its law is invariant under orthogonal conjugation and that its entries on and above the diagonal are independent in any fixed orthonormal basis. In particular, the unrevealed block is independent of the columns already read. Conditional on the past, the next query direction is fixed in the remaining subspace. In a basis starting with this direction, its diagonal entry is $N(0, 2/n)$, its remaining column entries are independent $N(0, 1/n)$, and the smaller block is independent of both. By induction, the assertion holds for adaptive queries, including conditioning on independent initial data. □

Lemma 13.2 (Operator norm)**.** *For every $n \geq 2$,*

$$\gamma_n\{g : \|W(g)\|_{\mathrm{op}} > 12\} \leq 2e^{-n}. \tag{13.5}$$

Proof. We use the sphere-net argument of [113, Corollary 4.2.13 and Exercise 4.4.3(b)]. Let $\mathcal{N}$ be a $1/4$ net of the Euclidean unit sphere with $|\mathcal{N}| \leq 9^n$, obtained by comparing the volumes of disjoint balls. For a symmetric matrix, approximating a unit vector at which the absolute value of the quadratic form is maximal by a point of the net, we have $\|W\|_{\mathrm{op}} \leq 2 \max_{u \in \mathcal{N}} |u^{\mathsf{T}} W u|$. For every fixed unit u, the quadratic form $u^{\mathsf{T}} W u$ is centered Gaussian with variance $2/n$. Using the Gaussian tail and a union bound, we find that

$$\gamma_n\{\|W\|_{\mathrm{op}} > 12\} \leq 2\, 9^n e^{-9n} = 2e^{(\log 9 - 9)n} \leq 2e^{-n}.$$

□

Lemma 13.3 (Pressure tail)**.** *Fix $\beta > 1$ and $0 < \mathrm{z_P} < 1/2$. There exists $c_\beta > 0$ such that*

$$\gamma_n\{F^c(g) > nP_\beta + n^{1-\mathrm{z_P}}\} \leq e^{-c_\beta n^{1-2\mathrm{z_P}}} = o(1). \tag{13.6}$$

Proof. Differentiating with respect to the primitive entries, we have

$$\partial_{g_{ij}} F^c(g) = \frac{\beta}{\sqrt{2n}} \sum_x \pi_g(x) x_i x_j, \qquad \|\nabla F^c(g)\|^2 \le \frac{\beta^2 n}{2}.$$

By Theorem 4.4 and Lemma 3.1(c), $\mathbb{E}_{\gamma_n} F^c = \mathbb{E} F_n \le n P_\beta$. Then (13.6) follows from Gaussian concentration [112], with the Lipschitz constant given by the gradient bound. □

13.3 The exponent parameters

Fix $\omega \in (0, 1/4)$ and define

$$\varepsilon_\omega = \frac{1/4 - \omega}{20} \in (0, 1/80). \tag{13.7}$$

We use the exponents in (13.8) throughout this Part. The strict inequalities between them are verified in Section 19.5, where we also give the instance $\omega = 1/5$. The letters r and w denote exponents, whereas $r(t)$ is a scalar profile and γ_n is Gaussian measure.

Exponent	Value	Exponent	Value
a	$1/2$	b	$1/2 - 14\varepsilon_\omega$
f	$1/2$	z	$1/4 - 8\varepsilon_\omega$
$\mathsf{k_f}$	$1/2 - 2\varepsilon_\omega$	w	$1/8 - 3\varepsilon_\omega$
D	$1/4 - \varepsilon_\omega$	$\mathsf{k_b}$	$1/4 - 9\varepsilon_\omega$
$\mathsf{u_E}$	$1/4 - 2\varepsilon_\omega$	$\mathsf{s_b}$	$1/4 - 5\varepsilon_\omega$
d	$\varepsilon_\omega/2$	$\mathsf{s_{seq}}$	$1/4 - 4\varepsilon_\omega$
e	$1/4 - 3\varepsilon_\omega$	r	$1/4 - 15\varepsilon_\omega/2$
$\mathsf{u_F}$	$1/4 - 39\varepsilon_\omega/4$	q	$1/4 - 10\varepsilon_\omega$
$\mathsf{z_P}$	$3/10$	$\mathsf{g_{rot}}$	$3/5$

(13.8)

13.4 The scalar Parisi process and its continuation

We use Theorems 4.4, 4.5, and 4.7, Proposition 4.3, and Corollary 4.8 and the scalar evolution identities of Section 5. Let μ_β be the unique minimizing Parisi measure and let $q_\beta \in (0,1)$ be its maximal support point. In variance time $v = \beta^2 t$, define

$$v_* = \beta^2 q_\beta, \qquad \bar\alpha(v) = \alpha_\beta(v/\beta^2), \qquad U(v,y) = \Phi_{\alpha_\beta}(v/\beta^2, y) \quad (0 \le v \le \beta^2).$$

Then $\partial_v U + U_{yy}/2 + \bar\alpha U_y^2/2 = 0$, with terminal value $\log(2\cosh y)$ at β^2. We continue the potential and the distribution function by

$$U(v,y) = \log(2\cosh y) + (\beta^2 - v)/2, \qquad \bar\alpha(v) = 1 \quad (v \ge v_*). \tag{13.9}$$

This agrees with the original solution on $[v_*, \beta^2]$ and satisfies the same equation for all later times. Let $m = U_y$ and $\chi = U_{yy}$. Up to β^2, these are $u_{\alpha_\beta}(v/\beta^2, y)$ and $\chi_{\alpha_\beta}(v/\beta^2, y)$, and for $v \geq v_*$ they are $\tanh y$ and $\operatorname{sech}^2 y$. We write $b(v, y) = \bar{\alpha}(v) m(v, y)$ before v_* and $b = m$ afterward. The value of the distribution function at a time atom does not affect its time integrals.

On canonical Wiener space with driver M, let

$$Y_t = M_t + \int_0^t b(v, Y_v)\, \mathrm{d}v, \qquad m_t = m(t, Y_t), \qquad \chi_t = \chi(t, Y_t). \tag{13.10}$$

Here and below, the time argument of Y is variance time. Up to β^2, the dictionary to (4.6) is $Y_v^{\mathrm{IV}} = Y_{v/\beta^2}$ and $M_v = \beta B_{v/\beta^2}$. Since the drift is bounded and spatially 1-Lipschitz, there exists a unique continuous, measurable, causal Picard solution. We have

$$m_0 = 0, \qquad |m(t, y)| < 1, \qquad 0 \leq \chi(t, y) \leq 1 - m(t, y)^2 \leq 1. \tag{13.11}$$

The potential and χ are even in y, and m and b are odd. These bounds follow from Proposition 4.3 and the step-recursion bound in [85, Section 3]. For the latter bound, we note that one Gaussian step at mass $\lambda \in [0, 1]$ satisfies $(Tf)' = \mathbb{E}_\lambda f'$ and $(Tf)'' = \mathbb{E}_\lambda f'' + \lambda \operatorname{Var}_\lambda(f')$, where $\mathbb{E}_\lambda$ is the normalized tilted expectation. If $0 < f'' \leq 1 - (f')^2$, then $(Tf)'' \leq 1 - ((Tf)')^2$. By induction from $\log(2\cosh y)$ and scalar stability, we obtain the inequality for χ. Since $\chi > 0$ by Proposition 4.3(i), this implies that $|m| < 1$. On the continuation, the formulas are explicit.

Proposition 13.4 (Scalar foundation)**.** *Under the Wiener law in* (13.10)*, m_t is a bounded martingale in the whole Y-past. Define*

$$\begin{gathered} p(t) = \mathbb{E}\chi_t, \qquad r(t) = \mathbb{E}\chi_t^2, \qquad \varkappa(t) = \mathbb{E}\frac{\chi_t^2}{1 - m_t^2}, \\ \rho(t) = \beta\sqrt{r(t)}, \qquad \sigma(t) = \sqrt{1 - \rho(t)^2}. \end{gathered} \tag{13.12}$$

All expectations are under the unlinked, zero-seed scalar Wiener law. For $0 \leq s \leq t$, we have

$$\mathbb{E}(m_t - m_s)^2 = \int_s^t r(v)\, \mathrm{d}v, \qquad \int_0^t r(v)\, \mathrm{d}v = \mathbb{E}m_t^2 \leq 1, \tag{13.13}$$

$$r = \beta^{-2}, \quad \rho = 1, \quad \sigma = 0 \quad (t \leq v_*), \qquad 0 < r(t) \leq \beta^{-2}, \quad p(t)^2 \leq r(t), \tag{13.14}$$

$$r \text{ is nonincreasing}, \quad |r'| \leq 2r, \qquad p \text{ is nonincreasing}, \quad |p'| \leq p \quad \text{almost everywhere}. \tag{13.15}$$

After v_*, we have $p = 1 - \mathbb{E}m_t^2$, $\varkappa = p$, and $p' = -r$. The functions ρ and σ are continuous and of bounded variation on every compact interval. Define

$$e(V) = \beta \int_0^V r(t)p(t)\, \mathrm{d}t, \qquad j(V) = \frac{1}{2}\int_0^V \rho(t)^2 \varkappa(t)\, \mathrm{d}t. \tag{13.16}$$

For $V \geq v_*$,

$$j(V) - \log 2 + P_\beta - \beta e(V) = \frac{\beta^2}{4} p(V)^2, \qquad \mathbb{E}_{\gamma_n} F^c \leq n P_\beta. \tag{13.17}$$

For a fresh solution started at $Y_s = y$, we also have

$$\mathbb{E}[Y_t m(t, Y_t)] - y m(s, y) = \int_s^t \mathbb{E}[\chi(v, Y_v) + m(v, Y_v) b(v, Y_v)] \, \mathrm{d}v. \tag{13.18}$$

This holds separately before and after v_*, and its integrand is identically one after v_*.

Proof. We begin with the martingale and profile identities, which follow from the scalar inputs. Then we identify the pressure normalization in the entropy–energy identity and prove the restart formula.

By Proposition 4.3(iv), with $v = \beta^2 t$, the process m is a bounded martingale before v_*. The explicit continuation has the same property by Itô's formula. The variance identity (13.13) follows from the scalar evolution in Proposition 4.3(iv) and the continuation. By Corollary 4.8, $\mathbb{E} m_v^2 = v/\beta^2$ on $[0, v_*]$. Continuity implies that $r(v) = \beta^{-2}$ on that closed interval. Using Jensen's inequality and (13.11), we find that $p^2 \leq r \leq p$.

For $0 \leq v \leq v_*$, the identity (4.9) and $\Gamma(v/\beta^2) = v/\beta^2$ imply

$$p(v) = 1 - \int \max(q, v/\beta^2) \, \mu_\beta(\, \mathrm{d}q). \tag{13.19}$$

Differentiating this Lipschitz integral almost everywhere, we obtain $p' = -\bar{\alpha}/\beta^2 = -\bar{\alpha} r$. This also follows directly from Proposition 4.3(iv). After v_*, we have $\chi_t = 1 - m_t^2$, so that $p = 1 - \mathbb{E} m_t^2$, $\varkappa = p$, and $p' = -r$. In both intervals, $|p'| \leq p$ because $r \leq p$. By the post-support estimate in [85, Lemma 3], r is nonincreasing and $-2r \leq r' \leq 0$; before v_* it is constant. This proves (13.14) and (13.15). Since r is continuous and monotone, the functions ρ and σ are continuous and have bounded variation.

In variance time, Theorems 4.4 and 4.5 identify

$$P_\beta = U(0,0) - \frac{1}{2\beta^2} \int_0^{\beta^2} v \bar{\alpha}(v) \, \mathrm{d}v.$$

This is the counting-measure normalization used in [85, Lemma 4]. The identity there reads $\log 2 - j(V) + \beta e(V) = P_\beta - \beta^2 p(V)^2/4$, which is (13.17); the endpoint $V = v_*$ follows by continuity. The pressure inequality follows from Theorem 4.4 and Lemma 3.1(c).

For the restart identity, the scalar evolution of Proposition 4.3(v), after the time change, gives zero drift for $m(v, Y_v)$. By the product rule [64, Theorem 3.3.6], the drift of $Y_v m(v, Y_v)$ is $\chi(v, Y_v) + m(v, Y_v) b(v, Y_v)$. We use localization and the bounded spatial derivatives of Section 5 to justify the product rule under step approximation. The stochastic integrand is $Y_v \chi(v, Y_v) + m(v, Y_v)$, which is square integrable on every finite interval because b is bounded. The drift is bounded by two, so we may take expectations and pass to the limit. This proves (13.18). On the continuation, its integrand is $\operatorname{sech}^2 Y_v + \tanh^2 Y_v = 1$. □

Lemma 13.5 (Cross moments and an integral representation). *For $0 \le s \le t \le H$, a bounded measurable test A of the Brownian past through s, and deterministic $f \in L^2([s,t])$,*

$$\mathbb{E}\left[A m_H \int_s^t f(v)\,\mathrm{d}M_v\right] = \int_s^t f(v)\mathbb{E}[A\chi_v]\,\mathrm{d}v. \tag{13.20}$$

Moreover, $m_t - m_s = \int_s^t \chi_v\,\mathrm{d}M_v$, with the integral constructed by predictable left sums in L^2 under the original Wiener law. These assertions hold across v_*.

Proof. We begin with the tested identity for Brownian increments, which we prove using the restart formula. We then pass to deterministic L^2 weights and identify the integral by its second moment.

First let A be a bounded Borel test of $Y_{[0,s]}$. The restart law is the product of its Y-past law with a fresh Wiener driver, followed by Picard restart. Multiplying (13.18) by A and integrating against this product law, we have

$$\mathbb{E}[A(Y_t m_t - Y_s m_s)] = \int_s^t \mathbb{E}[A(\chi_v + m_v b_v)]\,\mathrm{d}v, \qquad b_v = b(v, Y_v).$$

For unbounded endpoint observables, we clip the y factor of $ym(t,y)$ at $[-L,L]$ and take $L \to \infty$. The conditional first moments are dominated by $|Y_s| + \sqrt{t-s} + t - s$, which is integrable since $\mathbb{E}|Y_s| \le \sqrt{s} + s$. The drift $\chi + mb$ is bounded by two, so Fubini's theorem applies to the signed time integral. If the interval crosses v_*, we split there and retain the original past test on the second interval. By continuity, the split produces no endpoint term.

The martingale identity extends to integrable past tests by clipping, with domination by twice their absolute value. In particular, $\mathbb{E}[AY_s(m_t - m_s)] = 0$, and for each $v \in [s,t]$, $\mathbb{E}[Ab_v m_t] = \mathbb{E}[Ab_v m_v]$. Substituting $M_t - M_s = Y_t - Y_s - \int_s^t b_v\,\mathrm{d}v$, we obtain

$$\mathbb{E}[Am_t(M_t - M_s)] = \int_s^t \mathbb{E}[A\chi_v]\,\mathrm{d}v. \tag{13.21}$$

After completion, the sigma algebras of the Brownian past and of the field past agree: the Picard solution is causal, and $M_u = Y_u - \int_0^u b(v, Y_v)\,\mathrm{d}v$ recovers the driver causally. The recovery map into continuous paths is Borel by bounded joint measurability and rational evaluations. Thus (13.21) holds for Brownian-past tests. Projecting m_H to an intermediate time and using linearity, we obtain (13.20) for step functions f. By Brownian isometry and Cauchy–Schwarz, we may pass to L^2 weights: the two errors are at most $\|A\|_\infty \|f - f_\ell\|_2$ and $\|A\|_\infty \sqrt{t-s}\,\|f - f_\ell\|_2$.

For a refining deterministic partition $s = v_0 < \cdots < v_\ell = t$, let $Z_\ell = \sum_i \chi_{v_i}(M_{v_{i+1}} - M_{v_i})$. Predictable Gaussian isometry yields $\mathbb{E}Z_\ell^2 \to \int_s^t r(v)\,\mathrm{d}v$. Using (13.21) on each cell with past test χ_{v_i}, and projecting m_t to that cell's endpoint, we also have $\mathbb{E}[(m_t - m_s)Z_\ell] \to \int_s^t r(v)\,\mathrm{d}v$. In these limits, we used that χ is continuous along the paths and $\chi \le 1$. By (13.13), $\mathbb{E}|Z_\ell - (m_t - m_s)|^2 \to 0$, which proves the representation. □

On a pair $(M, \widetilde{M})$ of independent Wiener processes, we define

$$B = \int \rho \, \mathrm{d}M + \int \sigma \, \mathrm{d}\widetilde{M}, \qquad D = \int \sigma \, \mathrm{d}M - \int \rho \, \mathrm{d}\widetilde{M}. \tag{13.22}$$

For a deterministic continuous function f of bounded variation, the integral is the path function $I_f(b)(t) = f(t)b(t) - \int_{(0,t]} b(v) \, \mathrm{d}f(v)$ on paths starting at zero. By step approximation and Gaussian isometry, it agrees with the L^2 integral. The reflection matrix $\left(\begin{smallmatrix} \rho & \sigma \\ \sigma & -\rho \end{smallmatrix}\right)$ is symmetric and orthogonal at each time. It follows that (B, D) has the product Wiener law and has the causal inverse

$$M = \int \rho \, \mathrm{d}B + \int \sigma \, \mathrm{d}D, \qquad \widetilde{M} = \int \sigma \, \mathrm{d}B - \int \rho \, \mathrm{d}D.$$

These identities hold for the laws of the full pairs of paths and for the inverse maps. For step weights, the reflection squares to the identity on every independent Gaussian increment. Passing to uniform limits of paths, with integration by parts for functions of bounded variation, we find that the identities hold for ρ and σ. Taking a countable intersection over rational times, we identify the continuous versions. The future increments of (B, D) are independent of the pasts of both original drivers, since the approximating deterministic integrals use only fresh increments of $(M, \widetilde{M})$.

Let $\mathcal{W}$ denote Wiener measure. For a terminal sign x, we define the one-row law by

$$\nu_V(\mathrm{d}x, \mathrm{d}M, \mathrm{d}\widetilde{M}) = (1 + x m_V)(\mathrm{Unif}_1 \otimes \mathcal{W} \otimes \mathcal{W})(\mathrm{d}x, \mathrm{d}M, \mathrm{d}\widetilde{M}). \tag{13.23}$$

Averaging over the spin for each pair of paths, we see that this law is normalized and that the law of the entire pair of paths is unchanged. By martingale projection, we conclude that $\mathbb{E}_{\nu_V}[x \mid M_{[0,t]}, \widetilde{M}_{[0,t]}] = m_t$ for $t \le V$. We also use ν_V for its image on the history (x, Y, B, D) under the Picard solution map and (13.22). For $0 \le s \le t \le V$ and deterministic $f \in L^2([s,t])$, by Lemma 13.5 and the independence of the second original driver, we have

$$\begin{aligned} \mathbb{E}_{\nu_V}\left[x \int_s^t f(v) \, \mathrm{d}B_v\right] &= \int_s^t f(v)\rho(v)p(v) \, \mathrm{d}v, \\ \mathbb{E}_{\nu_V}\left[x \int_s^t f(v) \, \mathrm{d}D_v\right] &= \int_s^t f(v)\sigma(v)p(v) \, \mathrm{d}v. \end{aligned} \tag{13.24}$$

The tested conditional form of the first identity holds with any bounded test A of both pasts at s, with right side $\int_s^t f(v)\rho(v)\mathbb{E}[A\chi_v] \, \mathrm{d}v$. For the M term, we first integrate the independent $\widetilde{M}$ past; for the $\widetilde{M}$ term, we first integrate its centered fresh future. Unbounded Brownian tests are handled by clipping in L^2. Under ν_V, the path marginal remains Wiener, but adjoining the terminal sign to the filtration need not preserve the Brownian property. The filtrations used later are recorded in Section 14.2.

Lemma 13.6 (Terminal and inverse-profile bounds)**.** *There exist constants depending only on β such that, for all $t \geq 0$,*

$$\begin{aligned} p(t) &\leq C_\beta \frac{e^{-t/2}}{\sqrt{1+t}}, \qquad r(t) \geq c_\beta \frac{e^{-t/2}}{\sqrt{1+t}}, \\ R_V &= \left(\min_{0 \leq t \leq V} r(t) \right)^{-1} \leq C_\beta e^{V/2} \sqrt{1+V}. \end{aligned} \tag{13.25}$$

In particular, $p(V)^2 \leq C_\beta e^{-V}$.

Proof. By [85, Lemma 3, proof], the post-support transition kernel is

$$K_t(y,z) = e^{-t/2} \varphi_t(z-y) \frac{\cosh z}{\cosh y},$$

where φ_t is the centered Gaussian density of variance t. It integrates to one, and direct differentiation gives the generator $\partial_{zz}/2 + \tanh(z)\partial_z$. Since the drift is bounded and Lipschitz, uniqueness of the solution of the Picard equation identifies its law. Let λ_* be the endpoint law at v_*. For $k > 1/2$, we have

$$\mathbb{E}\chi^k_{v_*+t} = e^{-t/2} \int \lambda_*(\,\mathrm{d}y) \operatorname{sech} y \int \varphi_t(z-y) \operatorname{sech}^{2k-1} z \,\mathrm{d}z.$$

For $k = 1$ and $t \geq 1$, the bounds $\varphi_t \leq (2\pi t)^{-1/2}$ and $\int \operatorname{sech} z \,\mathrm{d}z = \pi$ imply the upper bound. For $k = 2$, let $C'_\beta < \infty$ be such that $\lambda_*([-C'_\beta, C'_\beta]) > 0$, and restrict the inner integral to $|z| \leq 1$. For $t \geq 1$ and $|y| \leq C'_\beta$, we have

$$\varphi_t(z-y) \geq \frac{e^{-(C'_\beta+1)^2/2}}{\sqrt{2\pi t}},$$

which implies the lower bound. On $0 \leq t \leq 1$, (13.15) implies that $r(v_* + t) \geq \beta^{-2} e^{-2t}$. The finite initial interval is absorbed into the constants, which depend only on the fixed temperature. The inverse bound follows from the monotonicity of r. □

SECTION 14

The Gaussian-query proposal on a growing mesh

In this section, we construct the proposal used in Sections 15, 17, and 18. We first prove the Gaussian assembly and reconstruction identities in Lemma 14.1. We then identify the law of the complete histories of the rows and prove the quadrature estimate in Lemma 14.3. Throughout, we use the scalar inputs of Section 13.4.

We follow the Gaussian-conditioning construction of [85, Section 4]. Successive Gaussian revelations along orthogonalized directions occur in the iterative construction of Bolthausen [25, Sections 2–3]. Montanari used related increments guided by the Parisi process in an incremental algorithm for approximate message passing [75, Sections 2 and 3.1]. Here the durations of the cells grow with time, and the source and replay constructions retain all the auxiliary coordinates needed for the row identities.

14.1 Growing schedule and source law

For $T > 0$, let $\mathcal{C}_T = C([0,T],\mathbb{R})$. We use the path-space conventions of Section 13. We write Br_h for the law of a Brownian bridge of duration $h > 0$, which is the image of Wiener measure under $b(u) \mapsto b(u) - (u/h)b(h)$. Fix n, $h > 0$, and an integer $N < n$. We define

$$t_0 = 0, \qquad h_j = he^{t_j/4}, \qquad t_{j+1} = t_j + h_j, \qquad V = t_{N+1}. \tag{14.1}$$

Cell zero is auxiliary. Query $j \in \{1,\dots,N\}$ uses

$$\begin{aligned} s_j^2 &= \int_{t_{j-1}}^{t_j} r(v)\,\mathrm{d}v, & \mathcal{T}_{j,i} &= \frac{m_i(t_j) - m_i(t_{j-1})}{s_j}, \\ b_j &= \frac{1}{\sqrt{h_j}} \int_{t_j}^{t_{j+1}} \rho(v)p(v)\,\mathrm{d}v. \end{aligned} \tag{14.2}$$

Here $m_i(t) = m(t, Y_i(t))$. The scalar reference b_j should not be confused with the drift $b(v,y)$ of the scalar equation. We use the normalized inner product and norm

$$\langle u, v\rangle = n^{-1}u^{\mathsf{T}}v, \qquad \|u\|^2 = \langle u, u\rangle.$$

We define the unlinked source law μ_0^{src} as the independent product of the laws of the following coordinates:

- n Wiener paths D_i through V and auxiliary Wiener paths B_i^0 of duration h_0;
- bridges $\mathrm{br}_{j,i} \sim \mathrm{Br}_{h_j}$ for $1 \le j \le N$ and $1 \le i \le n$;
- independent $e_j, v_j \sim N(0, I_n)$ and $d_j \sim N(0, 2/n)$ for $1 \le j \le N$;
- a full primitive $g^0 \sim \gamma_n$, with ambient matrix $W_0 = W(g^0)$, and a spin $x \sim \mathrm{Unif}_n$.

Unused bridges may be retained as independent ancillary coordinates; the estimates below do not depend on them. We write Ω^{src} for this product space. When we use primitive disorder, we adjoin the antisymmetric variable specified below, independently of the other inputs.

On the auxiliary cell, $B_i = B_i^0$. For $j \ge 1$, we append cell j by

$$B_i(t_j + u) = B_i(t_j) + \frac{u}{\sqrt{h_j}} e_{j,i} + \mathrm{br}_{j,i}(u), \qquad 0 \le u \le h_j. \tag{14.3}$$

Almost surely, and simultaneously for all the cells and rows used here, a bridge is unchanged when we subtract the line through its two endpoints. We define $M_i = \int \rho \, \mathrm{d}B_i + \int \sigma \, \mathrm{d}D_i$ and solve the scalar equation of Section 13.4 in each coordinate. Since the weights in these integrals are deterministic and of bounded variation, the integrals have the causal path interpretation given there. In particular, we can evaluate the targets in (14.2) before we append (14.3).

Applying the Gram–Schmidt procedure to the $\mathcal{T}_j$ in chronological order, we obtain directions u_j such that $\|u_j\| = 1$; when a residual vanishes, the choice of direction is deterministic and measurable. More precisely, let E have rows that are Euclidean orthonormal and span the remaining subspace. If $E\mathcal{T}_j \ne 0$, we normalize $E\mathcal{T}_j$; otherwise, we normalize the first nonzero column of E in the fixed coordinate order. Let q be the resulting vector, which has Euclidean norm one; the query direction is then $u_j = \sqrt{n} E^{\mathsf{T}} q$. We complete q to an orthonormal basis by a fixed measurable rule, rotate E by this basis, and retain all but its first row as the next remaining frame. We do this also when a residual vanishes. We define

$$U_j = \frac{1}{n} \sum_{l \le j} u_l u_l^{\mathsf{T}}, \qquad P_j = I - U_j, \qquad f_j = P_j e_j.$$

The assembled matrix and the completion vectors are defined by

$$\begin{aligned} W &= \frac{1}{n} \sum_{j=1}^{N} (u_j f_j^{\mathsf{T}} + f_j u_j^{\mathsf{T}}) + \frac{1}{n} \sum_{j=1}^{N} d_j u_j u_j^{\mathsf{T}} + P_N W_0 P_N, \\ C_j &= U_j e_j + P_j v_j. \end{aligned} \tag{14.4}$$

Note that the scalar recursion uses none of d_j, v_j, W_0, or x.

Lemma 14.1 (Gaussian assembly, completion, and path reconstruction)**.** *Under μ_0^{src}, conditional on all initial data (D, B^0, br), the joint law of $(W, (C_j)_j)$ is $\mathcal{G}_n \otimes N(0, I_n)^{\otimes N}$. This law does not depend on the initial data. The fixed-matrix replay endpoint*

$$e_j^{\mathrm{rep}} = P_j W u_j + U_j C_j \tag{14.5}$$

equals e_j on the source realization. The reconstructed (B_i, D_i) have the product of n independent Wiener-pair laws through V.

Proof. We first prove the conditional Gaussian law by induction on the remaining queries. We then expand the assembled matrix to identify the replay endpoints. Finally, we compute the covariance of each appended path to show that the reconstructed pairs have the Wiener law.

Conditional Gaussian law. For a fixed remaining frame, the column has Gaussian coordinates, consisting of a diagonal $N(0, 2/n)$ entry and independent off-diagonal $N(0, 1/n)$ entries. The projection $P_j e_j$ gives the off-diagonal column: when we multiply a Euclidean unit direction by $\sqrt{n}$ to obtain u_j, this cancels the $1/\sqrt{n}$ scale of the column. Conditional on the previously read e's and the initial data, the next frame is fixed.

We decompose the fresh e_j, v_j along U_j, P_j. Since the projections are orthogonal, the four Gaussian components are independent. It follows that $U_j e_j + P_j v_j$ is standard Gaussian and independent of $P_j e_j$ and d_j. By backward induction over the remaining queries, we find that the unrevealed block is a GOE compression on the remaining subspace, with the same $1/n$ normalization of the variance. It is independent of the part already read and of the later completion outputs. The base case is given by the independent ambient matrix W_0. Using orthogonal invariance as in Section 13.2, we assemble each column with this block. This proves the assertion on the joint law, including all later C_j, and that this law does not depend on the initial data.

Replay endpoints. Expanding (14.4), we find $P_j W u_j = P_j e_j$. Indeed, the earlier terms lie in the range of U_j, and the later terms and the final compression annihilate u_j. We also have $U_j C_j = U_j e_j$, so (14.5) equals e_j. By induction, we conclude that every replay target and frame agrees pointwise with the source, as does every initial segment of the scalar paths.

Pair paths. On one cell, $(u/\sqrt{h})e + \mathrm{br}(u)$ is a centered continuous Gaussian path, and its covariance at u, v is

$$\frac{uv}{h} + \min(u, v) - \frac{uv}{h} = \min(u, v).$$

So the path has the Wiener law on this cell. Concatenating independently over the deterministic cells, we obtain jointly Gaussian evaluations at any finite set of times, with covariance $\min(s, t)$, including across cells. Since evaluations at rational times determine

a Borel law on continuous paths, this identifies the law of the concatenation. The resulting pairs are independent across rows, and the B paths are independent of the D paths. Since (14.3) preserves every earlier initial segment, we find by induction that the concatenation is the reconstructed terminal state. The assertions also hold when unused bridges are retained. □

14.2 Spin laws, histories, and filtrations

Definition 14.2 (The two spin laws and the replay space)**.** On the source, we define

$$L = \prod_{i=1}^{n} (1 + x_i m_i(V)), \qquad \Lambda^{\mathrm{src}} = L \mu_0^{\mathrm{src}}. \tag{14.6}$$

The unlinked replay law μ_0^{rep} is the product of $\mathcal{G}_n$, independent $C_j \sim N(0, I_n)$, all initial B^0 paths and bridges, the n independent D paths, and Unif_n. Its paths are calculated by (14.5) and (14.3). We define $\Lambda^{\mathrm{rep}} = L^{\mathrm{rep}} \mu_0^{\mathrm{rep}}$ using the terminal fields in replay. We write Ω^{rep} for this product space, with the independent antisymmetric variable adjoined when using primitive disorder. The output law Q_V is the (g, x) marginal of Λ^{rep} on this primitive lift.

For each fixed choice of all inputs except the spin, we have $2^{-n} \sum_x L = 1$. This implies that both linked laws are probability measures, and that linking does not change the marginals obtained by forgetting the spin, including the marginal of the Gaussian matrix. Conditional on the paths, the terminal spins are independent with masses $(1 + x_i m_i(V))/2$. Equivalently, we may draw the spins using independent uniform $[0, 1]$ thresholds, with the same integrals in either realization.

Let $\Omega_V^{\mathrm{row}} = \{-1, 1\} \times \mathcal{C}_V^3$ be the canonical Borel space of histories of (x, Y, B, D) for one row through V. We continue to write ν_V for the one-row law of Section 13.4 mapped into this space by the reflected pair and the Picard solution. From $(\Omega_V^{\mathrm{row}})^n$ and independent ancillary coordinates, we reconstruct the source as follows. We restrict B to cell zero and recover, for $j \geq 1$,

$$\begin{aligned} e_{j,i} &= \frac{B_i(t_{j+1}) - B_i(t_j)}{\sqrt{h_j}}, \\ \mathrm{br}_{j,i}(u) &= B_i(t_j + u) - B_i(t_j) - \frac{u}{h_j} \big(B_i(t_{j+1}) - B_i(t_j) \big). \end{aligned} \tag{14.7}$$

The ancillary coordinates are d_j, v_j, g^0, the antisymmetric variable A used in the final lift, and the unused bridges; they are independent, with law

$$\begin{aligned} \mu^{\mathrm{anc}} = N(0, 2/n)^{\otimes N} \otimes N(0, I_n)^{\otimes N} \otimes \gamma_n \otimes \gamma_n^{\mathrm{anti}} \\ \otimes \bigotimes_{j \notin \{1, \dots, N\}} \mathrm{Br}_{h_j}^{\otimes n}. \end{aligned} \tag{14.8}$$

The last factor is omitted if unused bridges are not retained, and γ_n^{anti} is specified below. We write $\mathcal{A}$ for the ancillary product space. Since $W_0 = W(g^0)$, we retain the whole primitive matrix g^0, including its unused antisymmetric coordinates.

The endpoints and bridges in (14.7) are independent, since their evaluations at any finite set of times are jointly Gaussian with zero cross covariance. We have therefore identified the product law on the source space. Using the causal inverse of the reflected pair in Section 13.4, we identify the innovation and the scalar field in this reconstruction. All these maps are Borel under the conventions of Section 13. Indeed, evaluations at finitely many times, restrictions of paths, integration against weights of bounded variation, and the Picard solution map are Borel, and the frame rule uses finitely many measurable zero tests and normalizations.

The fixed-matrix experiment and its independent coordinate-path law are those of [85, Section 4.1]. The source and replay maps below retain the full histories and completion variables on the growing mesh, making their equality of laws explicit. We record the laws in the following diagram, in which every arrow except the density arrow denotes an equality of laws under a measurable pushforward. The same top arrows, applied before linking, identify the unlinked laws. The map F retains the initial data and D and replaces the source noises by (W, C) from (14.4).

$$\begin{array}{ccc} \big((\Omega_V^{\mathrm{row}})^n \times \mathcal{A}, \nu_V^{\otimes n} \otimes \mu^{\mathrm{anc}}\big) & \xrightarrow{\mathrm{Rec}} & (\Omega^{\mathrm{src}}, \Lambda^{\mathrm{src}}) \xrightarrow{F} (\Omega^{\mathrm{rep}}, \Lambda^{\mathrm{rep}}) \\ (\Omega^{\mathrm{rep}}, R_D = E_D \mu_0^{\mathrm{rep}}) & \xrightarrow{d\Lambda^{\mathrm{rep}}/dR_D = L^{\mathrm{rep}}/E_D} & (\Omega^{\mathrm{rep}}, \Lambda^{\mathrm{rep}}) \\ \downarrow{\scriptstyle (g,x)} & & \downarrow{\scriptstyle (g,x)} \\ (\mathbb{R}^{n\times n} \times \Sigma_n, \gamma_n \otimes \mathrm{Unif}_n) & & Q_V. \end{array} \tag{14.9}$$

Here Rec is the reconstruction just described, using (14.7) and the reflected pair. Along the density arrow, we change measure on the same replay space; E_D and R_D are constructed in Section 15.1. In particular, the top row implies the identity for the complete histories

$$\big((x_i, Y_i, B_i, D_i)_{i=1}^n\big)_* \Lambda^{\mathrm{rep}} = \nu_V^{\otimes n}. \tag{14.10}$$

To check the linked identities, let ν_V^0 be the canonical history law before the one-row link. By the factorization into endpoints and bridges and the inverse of the reflection, we have

$$\mathrm{Rec}_*\big((\nu_V^0)^{\otimes n} \otimes \mu^{\mathrm{anc}}\big) = \mu_0^{\mathrm{src}}.$$

On a supporting set of full measure, the recomputed field agrees with the given Y. For every bounded Borel test function f on the source space, we have

$$\int f\big(\mathrm{Rec}(w,a)\big) \prod_{i=1}^n \big(1 + x_i m(V, Y_i(V))\big)\, \mathrm{d}\big((\nu_V^0)^{\otimes n} \otimes \mu^{\mathrm{anc}}\big)(w,a) = \int f L \,\mathrm{d}\mu_0^{\mathrm{src}}.$$

This gives the law on the source space in the top row, with every ancillary coordinate retained. Repeating this calculation of densities and maps for F, and using the pointwise replay identity in Lemma 14.1, we obtain the pushforward identity for the linked replay law. The matrix and completion vectors are computed from these rows and ancillary variables; they need not be independent of the rows.

For the primitive g in (14.9), we retain the antisymmetric Gaussian variable A, independently of all the other inputs. Its entries A_{ij}, $i < j$, are independent $N(0, 1/2)$, with $A_{ji} = -A_{ij}$ and $A_{ii} = 0$. We define

$$g = \sqrt{n/2}\, W + A.$$

Then $g \sim \gamma_n$ and $W(g) = W$, as in Section 13.1. Since replay reads only W, this gives the proposal with primitive input. The proposal and the target have the same conditional law of A given (W, x). Their joint relative entropy on primitive matrices equals their relative entropy on physical matrices. We use this lift whenever we condition on g.

Filtrations. Whether a Brownian assertion may be used depends on the law and on the information revealed at time zero. The following table records the distinctions needed in the later proofs. In the last row, $\xi_i(v) = x_i \chi(v, Y_i(v))/(1 + x_i m(v, Y_i(v)))$, as in Section 15.1.

Law and filtration	Brownian or independence assertion	Use
Scalar Wiener pair, with the spin-free pair past	$(M, \widetilde{M})$ and its deterministic orthogonal transform (B, D) are Wiener pairs.	Scalar martingales, brackets, and path estimates.
Unlinked row, with an independent uniform sign revealed initially	(B, D) remains a Wiener pair in the filtration including this sign.	Signed integrals and fixed-sign Itô identities.
Linked row ν_V, with the sign revealed initially	The entire path marginal is unchanged; B, D need not be Brownian in the filtration including the sign.	Posterior moments, path-only estimates, and the previously defined signed integrals.
Product of complete linked rows and independent ancillary variables	The complete rows are independent before conditioning on W.	Independent-row concentration estimates.
Unlinked replay, with $(W, C, B^0, \mathrm{br}, x)$ revealed initially	D is Brownian and independent of the static inputs. No Brownian property for B is asserted in this filtration.	Construction and normalization of R_D.
Linked source natural filtration, including x, excluding the final matrix and unused matrix noises	The compensated vectors $B - \int \rho\xi \, \mathrm{d}v$ and $D - \int \sigma\xi \, \mathrm{d}v$ are Brownian.	The mean information identity by bounded-integrand Girsanov.

Once the terminal sign or the assembled matrix is revealed, a Brownian assertion must be verified separately in the resulting filtration. We transport all scalar expectations at fixed V and all Gram tails below by (14.10). We use the bound $L \le 2^n$ only for null sets at fixed n or for integral versions, as in Section 13.

14.3 Population entries

These are the martingale and posterior-projection identities used in [85, Section 4, proof of Proposition 5]; we record their form for the unequal neighboring cell durations. For a single linked row, let $e_j = (B(t_{j+1}) - B(t_j))/\sqrt{h_j}$ and $\mathcal{T} = (\mathcal{T}_j)_{j=1}^N$, $e = (e_j)_{j=1}^N$. By Lemma 13.5, the identities for the reflected pair in Section 13.4, and posterior projection, we have

$$\begin{gathered} \mathbb{E}\mathcal{T}\mathcal{T}^{\mathsf{T}} = I_N, \qquad \mathbb{E}ee^{\mathsf{T}} = I_N, \qquad \mathbb{E}\mathcal{T}x = s, \qquad \mathbb{E}ex = b, \\ H_0 = \mathbb{E}\mathcal{T}e^{\mathsf{T}}, \qquad (H_0)_{ij} = 0 \quad (i \le j). \end{gathered} \tag{14.11}$$

Indeed, disjoint martingale increments are orthogonal. The identity $\mathbb{E}\mathcal{T}_j x = s_j$ follows from the variance identity in Proposition 13.4, and the reflected cross moment gives $\mathbb{E}e_j x = b_j$. For $i \le j$, $\mathcal{T}_i$ is measurable before the fresh B cell defining e_j, so the path-only moment is zero. We retain the entire lower block H_0. By Bessel's inequality in the L^2 space for one row, we have

$$\|s\|, \|b\|, \|H_0\| \le 1, \qquad 0 \le e_h = s^{\mathsf{T}} b \le 1. \tag{14.12}$$

The row means vanish by zero-field symmetry. Note that these calculations do not use independence of the entries within a row.

Lemma 14.3 (Adjacent-cell quadrature and stopping)**.** *If* $\max_{j \le N} h_j \le 1/8$, *then*

$$|e_h - e(V)| \le C_\beta h. \tag{14.13}$$

For $h = n^{-\mathsf{a}}$ *and* $\mathsf{L} > 0$, *define*

$$N_{\mathsf{L}} + 1 = \min\{j : t_j \ge \mathsf{L} \log n\}, \qquad V_{\mathsf{L}} = t_{N_{\mathsf{L}}+1}.$$

If $\mathsf{a} > \mathsf{L}/4$, *then eventually*

$$\begin{gathered} N_{\mathsf{L}} + 1 \le 8n^{\mathsf{a}}, \qquad \mathsf{L} \log n \le V_{\mathsf{L}} \le \mathsf{L} \log n + n^{-\mathsf{a}+\mathsf{L}/4}, \\ \max_{j < N_{\mathsf{L}}+1} h_j \le n^{-\mathsf{a}+\mathsf{L}/4}. \end{gathered} \tag{14.14}$$

There are $N_{\mathsf{L}} + 1$ *cells and exactly* N_{L} *matrix queries. If additionally* $\mathsf{a} < 1$, *then eventually* $N_{\mathsf{L}} + 1 \ge 2$ *and* $N_{\mathsf{L}} < n$.

Proof. We compare the normalization from the preceding cell with r on the next cell. Using the relative variation bounds of Proposition 13.4, we obtain an integrable quadrature error. For the stopping estimate, we sum the decrements of $e^{-t_j/4}$.

For $v \in [t_j, t_{j+1}]$ and $u \in [t_{j-1}, t_j]$, the relative variation bounds, monotonicity, and $h_j/h_{j-1} = e^{h_{j-1}/4}$ imply that

$$1 - h_j/8 \le \frac{s_j}{\sqrt{h_j r(v)}} \le 1 + 4h_j. \tag{14.15}$$

For the upper bound, we have $r(v) \ge (1 - 4h_j) r(u)$ and $(1 - 4h_j)^{-1/2} \le 1 + 4h_j$ for $h_j \le 1/8$. For the lower bound, we use $r(u) \ge r(v)$ and $e^{-h_{j-1}/8} \ge 1 - h_j/8$. It follows that

$$\left| \frac{s_j}{\sqrt{h_j}} - \sqrt{r(v)} \right| \le 4h_j \sqrt{r(v)}.$$

With $q_e(v) = \beta r(v)p(v)$, we obtain

$$|e_h - e(V)| \le h + 4h \int_0^V e^{v/4} q_e(v)\,\mathrm{d}v. \tag{14.16}$$

The first term h bounds the contribution of the omitted auxiliary cell. The integral is bounded uniformly in V. Indeed, on the finite interval before the endpoint of the support, we have $q_e \le \beta$, and after it, using $r \le p$ and Lemma 13.6, we find that the integrand decays exponentially. This proves (14.13), with, for example,

$$C_\beta = 1 + 4\beta \int_0^\infty e^{v/4} p(v) r(v)\,\mathrm{d}v < \infty.$$

For $h_j \le 1$, we have $1 - e^{-h_j/4} \ge h_j/8$, and hence

$$e^{-t_j/4} - e^{-t_{j+1}/4} \ge h/8.$$

Summing up to the first reaching time gives $(N_{\mathsf{L}} + 1)h \le 8$. Every left endpoint that is used lies below $\mathsf{L}\log n$, which proves the bounds on the number of cells and on the overshoot in (14.14). The reaching index exists because $t_j \ge jh$. Since $\mathsf{L}\log n > 0$, it is at least one. If additionally $\mathsf{a} < 1$, it is eventually at least two, because $t_1 = n^{-\mathsf{a}} \to 0 < \mathsf{L}\log n$ and $\mathsf{a} > \mathsf{L}/4 > 0$. Further, (14.14) implies that

$$N_{\mathsf{L}} < N_{\mathsf{L}} + 1 \le 8n^{\mathsf{a}} < n \qquad \text{once } n^{1-\mathsf{a}} > 8.$$

In every fine or coarse use, the fine-energy conditions in Proposition 15.5 supply $0 < \mathsf{k}_{\mathsf{f}} < 1 - \mathsf{a}$, so $\mathsf{a} < 1$; in particular, $\mathsf{a} = 1/2$ in the exponent table and the range family of Sections 13.3 and 19.5. We never use $\mathsf{L} = 0$. All scalar references are evaluated at V_{L}, including its overshoot. The estimate needs neither an extra cell nor a quadrature error bounded by the maximal duration of a cell. □

We use the same sequence (14.1) for both $\mathsf{L} = \mathsf{b}$ and $\mathsf{L} = \mathsf{f}$, where $0 < \mathsf{b} < \mathsf{f}$, and the two replays use identical chronological inputs through V_{b}. By induction on the queries, every scalar initial segment through V_{b} is the same in the two replays. The auxiliary cells agree. If the scalar paths agree through the preceding cell, then the targets agree, the same remaining frame and matrix give the same (14.5), and the same bridge appends the same next cell. This argument covers the deterministic fallback as well. At query j, the number of remaining queries is $N_{\mathsf{L}} - j$; this does not change the chronological schedule. The terminal spins drawn at the fine horizon are never fed back into the continuation.

SECTION 15

Forward entropy at the fine horizon

The goal of this section is to prove Proposition 15.5, which bounds the relative entropy of the fine proposal with respect to Π_n. We use the source and replay laws of Definition 14.2, the scalar identities of Proposition 13.4 and Lemma 13.5, and the quadrature estimate of Lemma 14.3. We use this bound at the two ends of the connecting sequence in Section 18. For the reference measure and the comparison of likelihoods, we follow [85, Section 4, Proposition 5]. We then estimate the expected energy directly, including the outcomes on which the query uses the deterministic fallback.

Throughout this section, $V = t_{N+1}$ is the terminal time of a proposal with $N < n$ queries. We write $m_i(t) = m(t, Y_i(t))$ and $\chi_i(t) = \chi(t, Y_i(t))$. Expectations with subscript Λ refer to the linked source or replay law, according to the space of the integrand; these laws agree on the complete row histories and on the output by Section 14.2.

15.1 The reference measure

On the unlinked replay space, we regard $(W, C, B^0, \mathrm{br}, x)$ as static. The D_i are independent Brownian motions, independent of these static data. The replay field at time t is a Borel function of these static data and $D_{[0,t]}$: each next query reads the field only at the left endpoint of its new cell. We define

$$\xi_i(t) = \frac{x_i \chi(t, Y_i(t))}{1 + x_i m(t, Y_i(t))}, \qquad |\xi_i(t)| \le 2. \tag{15.1}$$

The coefficient $\sigma(t)\xi_i(t)$ is continuous and predictable in this replay filtration. It vanishes before v_*; afterward it is $\sigma(t)(x_i - m_i(t))$. Under μ_0^{rep}, let

$$\begin{aligned} \mathscr{M}_D &= \sum_i \int_0^V \sigma(t)\xi_i(t)\,\mathrm{d}D_i(t), \\ \langle \mathscr{M}_D \rangle &= \sum_i \int_0^V \sigma(t)^2 \xi_i(t)^2\,\mathrm{d}t \le 4nV, \\ E_D &= e^{\mathscr{M}_D - \langle \mathscr{M}_D \rangle/2}, \qquad R_D = E_D \mu_0^{\mathrm{rep}}. \end{aligned} \tag{15.2}$$

We define $\mathscr{M}_D$ as the L^2 limit of the vector of predictable sums at the left endpoints of dyadic partitions. This definition, and not the normalization alone, is what identifies it

with the sum of the row integrals below. We choose versions according to the conventions of Section 13.

Lemma 15.1 (The reference measure and the row integrals)**.** *The measure R_D is a probability measure, preserves the complete static marginal, and has output marginal $\gamma_n \otimes \mathrm{Unif}_n$ on (g, x). For one unlinked row, let $\xi(t) = x\chi_t/(1 + xm_t)$ and*

$$\mathscr{M}_D^{\mathrm{row}} = \int_0^V \sigma(t)\xi(t)\,\mathrm{d}D(t),$$
$$\langle \mathscr{M}_D^{\mathrm{row}} \rangle = \int_0^V \sigma(t)^2\xi(t)^2\,\mathrm{d}t.$$

The integral denotes its unlinked L^2 left-sum version; we use the same Borel version under the equivalent linked row law. Moreover,

$$\frac{\mathrm{d}\Lambda^{\mathrm{rep}}}{\mathrm{d}R_D} = L/E_D, \qquad I^{(n)} = \log(L/E_D) = \sum_{i=1}^n I_{V,i},$$
$$I_V = \log(1 + xm_V) - \mathscr{M}_D^{\mathrm{row}} + \tfrac{1}{2}\langle \mathscr{M}_D^{\mathrm{row}} \rangle, \tag{15.3}$$

Here $I_{V,i}$ is I_V evaluated on row i. The row histories are independent with law ν_V of Section 13.4, so the summands are independent with common law $(I_V)_*\nu_V$. Finally,

$$\mathbb{E}_{\Lambda^{\mathrm{rep}}} I^{(n)} = nj(V). \tag{15.4}$$

Proof. We first normalize the exponential conditional on the static data. We then transport its defining left sums to the source row space and compute the mean in the source natural filtration.

Normalization. This is the conditional stochastic-exponential normalization used in [85, Section 4, proof of Proposition 5]; the bounded-integrand change of measure is Girsanov's theorem [64, Theorem 3.5.1]. Fix the static data and a deterministic grid for the left sums. The last increment of D is Gaussian and independent of the earlier ones, and integrating it out shows that its factor in the exponential, together with the compensator, has conditional mean one. Iterating over the increments, we find that the exponential of the left sums has mean one. Since the predictable coefficients are bounded and continuous, their left approximants converge in $L^2(\,\mathrm{d}t\;\mathrm{d}\mu_0^{\mathrm{rep}})$. By isometry, the sums then converge to the integral defining $\mathscr{M}_D$. The exponentials are uniformly integrable, because the same Gaussian identity at twice the coefficient bounds their second moments by e^{4nV}. Testing against a bounded function of the complete static data before taking the limit, we conclude that the limiting exponential has conditional mean one given these data. This proves that R_D is normalized and that it preserves the complete static marginal, including the output marginal.

With the Borel versions fixed above, and after a modification on a null set, E_D is positive and finite everywhere. Since $L > 0$, multiplying the densities gives $R_D \cdot (L/E_D) = \mu_0^{\mathrm{rep}} \cdot L = \Lambda^{\mathrm{rep}}$.

Identification of the row integrals. To identify the integral, we use the diagram of laws and the history identity in Section 14.2 for the unlinked laws. At every finite dyadic level, the pullback of the vector sum on the replay space equals the sum of the corresponding functions of the row histories (x, Y, B, D). On the unlinked row space, D is a Brownian motion in the filtration generated by the two original drivers and by the independent uniform sign, which is revealed at time zero. The coefficient $\sigma\xi_i$ from (15.1) is bounded and predictable, so the same sums converge in L^2 to its integral. The law identity and the Cauchy–Schwarz inequality for finite sums imply that the sum of the row limits converges in L^2 on the replay space. By uniqueness of the limit, $\mathscr{M}_D$ is the sum of the row integrals. By Fubini's theorem in the time variable, the brackets split in the same way. This proves (15.3) almost everywhere under the unlinked law. The linked and unlinked laws are equivalent, which implies that the identity also holds almost everywhere under the linked law. Here the transported variables are the complete histories, and the versions are fixed at the given n, V.

The mean. To compute the mean, we work in the natural filtration of the unlinked source, $\mathcal{F}_t = \sigma(x, (B_i, D_i)_{[0,t]} : i \le n)$, which does not contain W, d_j, and W_0. By Lemma 13.5, we have $\mathrm{d}m_i = \chi_i(\rho\,\mathrm{d}B_i + \sigma\,\mathrm{d}D_i)$. Using Itô's formula, localized where $1 + x_i m_i \ge 1/l$ and then letting $l \to \infty$, we find

$$\log L = \sum_i \int_0^V \xi_i(\rho\,\mathrm{d}B_i + \sigma\,\mathrm{d}D_i) - \tfrac{1}{2}\sum_i \int_0^V \xi_i^2\,\mathrm{d}t.$$

The localization can be removed because $|\xi_i| \le 2$ and the denominator is strictly positive on every finite continuous sample path. The integral against D is the pullback identified above, so

$$I^{(n)} = \sum_i \int_0^V \rho\xi_i\,\mathrm{d}B_i - \tfrac{1}{2}\sum_i \int_0^V \rho^2\xi_i^2\,\mathrm{d}t. \tag{15.5}$$

The process $L_t = \prod_i (1 + x_i m_i(t))$ is the positive bounded martingale that defines the terminal link, with $L_0 = 1$. The expectation in Novikov's condition is at most e^{2nV}. By Girsanov's theorem for bounded integrands [64, Theorem 3.5.1], under Λ^{src} the processes

$$B_i^{\Lambda} = B_i - \int_0^{\cdot} \rho\xi_i\,\mathrm{d}t, \qquad D_i^{\Lambda} = D_i - \int_0^{\cdot} \sigma\xi_i\,\mathrm{d}t$$

form a Brownian vector in this natural filtration. It follows from (15.5) that

$$\mathbb{E}_{\Lambda} I^{(n)} = \frac{1}{2}\sum_i \mathbb{E}_{\Lambda} \int_0^V \rho^2\xi_i^2\,\mathrm{d}t.$$

Given the past of the spin-free pair at time t, the linked spin probabilities are $(1 + xm_t)/2$, and therefore

$$\sum_{x=\pm 1} \frac{1 + xm_t}{2} \frac{\chi_t^2}{(1 + xm_t)^2} = \frac{\chi_t^2}{1 - m_t^2}.$$

The spin-free paths keep their original law. Since the integrands are bounded, we may exchange the expectation and the time integral. Using the definition of j in Proposition 13.4, we obtain (15.4).

Integrability. All signed quantities used here are integrable under the indicated laws. By isometry, we have $\mathscr{M}_D \in L^2(\mu_0^{\mathrm{rep}})$; also $\langle \mathscr{M}_D \rangle \le 4nV$. For fixed n, the bound $L \le 2^n$ transfers this L^2 integrability to Λ. Also, $\log L \in L^2(\Lambda)$, since $t(\log t)^2$ is bounded on $0 < t \le 2$ and all other spin factors average to one. After compensation in (15.5), we obtain the bound $\mathbb{E}_\Lambda (I^{(n)})^2 \le 8nV + 8n^2V^2$. The calculation with conditional exponentials above also proves the case of Girsanov's theorem with bounded integrands, by exponential tilting of finitely many Gaussian steps and passage to the limit in L^2. Throughout, the filtration does not contain the final matrix. □

Lemma 15.2 (Forward relative entropy)**.** *For the output Q_V of Definition 14.2,*

$$\mathrm{KL}(Q_V \| \Pi_n) \le nj(V) - n \log 2 + \mathbb{E}_{\gamma_n} F^c - \beta \mathbb{E}_\Lambda H_g^c(x). \tag{15.6}$$

Proof. We use the lifted-reference and entropy-contraction argument of [85, Section 4, proof of Proposition 5]. On the replay space, we define $\mathrm{d}\widehat{P} = 2^n e^{\beta H_g^c(x) - F^c(g)} \, \mathrm{d}R_D$. Its total mass is one, since R_D projects to $\gamma_n \otimes \mathrm{Unif}_n$; its output law is Π_n. Further,

$$\log \frac{\mathrm{d}\Lambda}{\mathrm{d}\widehat{P}} = I^{(n)} - n \log 2 + F^c(g) - \beta H_g^c(x).$$

Averaging pointwise over the spins, we see that the matrix marginal of the linked law is the GOE law. In particular, with $H_*(W) = \max_x |x^{\mathsf{T}} W x / 2|$,

$$\mathbb{E}_\Lambda H_*^2 \le \frac{n^2}{4} \mathbb{E}_{\mathcal{G}_n} \|W\|_F^2 = \frac{n^2(n+1)}{4}. \tag{15.7}$$

Also $|F^c| \le n \log 2 + |\beta| H_*$. These bounds and Lemma 15.1 justify the finite KL calculation. The data processing inequality for the output map [28, Chapter 4, Exercise 4.10] and (15.4) imply (15.6). □

15.2 Fine energy

We now estimate the expected energy, one query at a time. The decomposition into queried columns, diagonal innovations and an unrevealed GOE block is the energy decomposition in [85, Section 4, proof of Proposition 5]. The conditional expectation

calculation below also controls the fallback outcomes and the unequal cell durations. Let $\mathcal{F}_{t_j}^{\mathrm{paths}}$ be the natural filtration of the spin-free pair of drivers in the source, through time t_j. Under the linked law, the marginal law of these paths is unchanged. The conditional laws of the future rows and of the signs factor across coordinates, because the unlinked rows factor, their pasts factor, and the link is a product. We obtain their conditional densities by integrating the terminal link of each row over its independent future. In particular, the pairs $(e_{j,i}, x_i)$ are conditionally independent across i and satisfy

$$\begin{aligned} \mathbb{E}[e_{j,i} \mid \mathcal{F}_{t_j}^{\mathrm{paths}}] = 0, \qquad \mathbb{E}[x_i \mid \mathcal{F}_{t_j}^{\mathrm{paths}}] = m_i(t_j), \\ \psi_i = \mathbb{E}[e_{j,i} x_i \mid \mathcal{F}_{t_j}^{\mathrm{paths}}] = \frac{1}{\sqrt{h_j}} \int_{t_j}^{t_{j+1}} \rho(v) \mathbb{E}[\chi_i(v) \mid \mathcal{F}_{t_j}^{\mathrm{paths}}] \, \mathrm{d}v. \end{aligned} \tag{15.8}$$

The last identity is the cross-moment identity of Section 13.4 tested against functions of the past. Here we condition on the past of the original drivers, not on W. The Picard map with fresh future drivers gives a Borel version of ψ_i that depends only on the past of its own row, in accordance with the conventions on versions in Section 13. If $h_{j-1}, h_j \le 1/8$, then the relative variation estimates of Proposition 13.4 and the growth of the schedule imply that

$$0 \le \psi_i \le \beta \sqrt{h_j r(t_j)} \le 2\beta s_j, \qquad \mathbb{E}\overline{\psi} = b_j, \quad \overline{\psi} = n^{-1} \sum_i \psi_i. \tag{15.9}$$

The factor two accounts for the different durations of the previous and current cells.

In this section, we write $\varrho_j = \|P_{j-1} \mathcal{T}_j\|$ for the residual norm. Whether the query is normalized or uses the fallback, we have $\langle u_j, m(t_j) \rangle = s_j \varrho_j$, because $m(t_j) = \sum_{l \le j} s_l \mathcal{T}_l$ and each $\mathcal{T}_l$ belongs to the span of the first l queries. Also $P_j m(t_j) = 0$.

Lemma 15.3 (Exact one-query conditional identity)**.** *With $c_j = \langle u_j, x \rangle$,*

$$\begin{aligned} &\mathbb{E}[c_j \langle x, P_j e_j \rangle \mid \mathcal{F}_{t_j}^{\mathrm{paths}}] \\ &\qquad = s_j \varrho_j \frac{1}{n} \sum_i (P_j)_{ii} \psi_i - \frac{2}{n^2} \sum_i u_{j,i} (P_j)_{ii} \psi_i m_i(t_j). \end{aligned} \tag{15.10}$$

Proof. We expand $(n^{-1} \sum_i u_i x_i)(n^{-1} \sum_{k,l} x_k P_{kl} e_l)$. Since the row pairs are conditionally independent and $x_i^2 = 1$, the mixed third moment is

$$\mathbb{E}[x_i x_k e_l \mid \text{past}] = \mathbf{1}_{l=i} \psi_i m_k + \mathbf{1}_{l=k} m_i \psi_k - 2\mathbf{1}_{i=k=l} m_i \psi_i.$$

For $i = k$ it is zero by Gaussian centering. For $i \ne k$, it follows from independence, also when l differs from both indices. The first contraction contains $P_j m(t_j) = 0$, and the second contains $\langle u_j, m(t_j) \rangle = s_j \varrho_j$. In more detail, given the past of a row, a fresh seed for the row consists of the current B bridge, the future B path after that cell, the

future D increments, and a terminal uniform threshold. It is independent of the standard Gaussian endpoint $e_{j,i}$; these seeds and endpoints are independent across coordinates. Since the Picard solution is causal, the terminal sign is a Borel function of its own endpoint and seed. These conditional product laws justify the third-moment calculation above. All terms are integrable, since the signs are bounded and Gaussian first moments are finite. □

Lemma 15.4 (Expected energy at the fine horizon)**.** *For $V = t_{N+1}$, $\max h_j \le 1/8$,*

$$\mathbb{E}_\Lambda \frac{H_g^c(x)}{n} \ge e(V) - C_\beta \left[h + \frac{(N+1)(1+V) + \sqrt{N}}{n} + \frac{(N+1)^2 V R_V}{n^2} \right]. \tag{15.11}$$

Proof. We first estimate the expectation in Lemma 15.3, including the residual norm and the projection diagonals. We then sum over the queries and use the quadrature estimate of Lemma 14.3.

The residual norm. Since $\sum_i |u_{j,i}| \le n$, the correction in (15.10) is at most $4\beta s_j/n$ in absolute value. Removing the projection diagonals from its first term costs at most $2\beta s_j^2 j\varrho_j/n$, and $\mathbb{E}\varrho_j \le \sqrt{\mathbb{E}\|\mathcal{T}_j\|^2} = 1$. For this calculation, we write $Q = \|\mathcal{T}_j\|^2$, $Y = \|U_{j-1}\mathcal{T}_j\|^2$, $X = \varrho_j^2 = Q - Y$. For $X \ge 0$,

$$\sqrt{X} \ge 1 + (X-1)/2 - (X-1)^2/2. \tag{15.12}$$

Indeed, the remainder in the tangent-line identity for the square root is $(\sqrt{X}-1)^2/2 \le (X-1)^2/2$. Since the row pairs $(\psi_i, \mathcal{T}_{j,i})$ are independent and identically distributed, we have

$$\mathbb{E}[\overline{\psi}(Q-1)] = n^{-1}\operatorname{Cov}(\psi_i, \mathcal{T}_{j,i}^2) \ge -2\beta s_j/n. \tag{15.13}$$

Here we used that $\mathbb{E}\mathcal{T}_{j,i}^2 = 1$ and $\psi_i \ge 0$.

Let $h_- = h_{j-1}$. Conditional on the past of the original pair up to time t_{j-1}, the coordinates of the new vector $\mathcal{T}_j$ are independent, centered martingale increments, with conditional second moment at most h_-/s_j^2 and fourth moment at most $3h_-^2/s_j^4$. The bound on the fourth moment follows from Itô's formula, applied to the fourth power of an increment with coefficient $\chi \le 1$, and the bound on the second moment. The rank-$(j-1)$ projection U_{j-1} is measurable at this time. Expanding its quadratic form using conditional centering and independence, we obtain

$$\mathbb{E}Y \le \frac{jh_-}{ns_j^2}, \qquad \mathbb{E}Y^2 \le \frac{3(j^2+j)h_-^2}{n^2 s_j^4}. \tag{15.14}$$

For the unconditional fourth moment, we use the elementary bound $\mathbb{E}Z_T^4 \le 144\mathbb{E}\langle Z\rangle_T^2$ for a martingale starting at zero. Indeed, by Itô's formula, we have

$$\mathbb{E}Z_T^4 \le 6(\mathbb{E}\sup|Z|^4)^{1/2}(\mathbb{E}\langle Z\rangle_T^2)^{1/2}.$$

Doob's fourth-moment inequality then implies the stated bound with constant 144. Since $\langle \Delta m \rangle \le h_-$ and its expectation is s_j^2, we have

$$\mathbb{E}\mathcal{T}_{j,i}^4 \le 144 h_- / s_j^2, \qquad \mathbb{E}(Q-1)^2 \le 144 h_- / (n s_j^2). \tag{15.15}$$

All conditional statements here refer only to the pasts of the original drivers.

Summation of the query energies. We multiply (15.12) by $s_j \overline{\psi} \ge 0$ and use (15.9), (15.13)–(15.15), and $(X-1)^2 \le 2(Q-1)^2 + 2Y^2$. Inserting the result into (15.10), we obtain

$$\begin{aligned} \mathbb{E}[c_j \langle x, P_j e_j \rangle] \ge s_j b_j - C_\beta \Bigg[& \frac{(j+1)s_j^2 + s_j + (j+1)h_-}{n} \\ & + \frac{(j^2+j)h_-^2}{n^2 s_j^2} \Bigg]. \end{aligned} \tag{15.16}$$

This derivation also covers a vanishing residual, $\varrho_j = 0$. When summing over the queries, we use $\sum s_j^2 \le 1$, $\sum s_j \le \sqrt{N}$, $s_j^2 \ge h_- / R_V$, and $\sum h_- \le V$.

Expanding the assembled matrix of Section 14.1, we have, on every outcome of the source,

$$\begin{aligned} \frac{H_g^c(x)}{n} &= E_{\mathrm{qry}} + Z, \qquad E_{\mathrm{qry}} = \sum_j c_j \langle x, P_j e_j \rangle, \\ Z &= \frac{1}{2} \sum_j c_j^2 d_j + \frac{1}{2n} (P_N x)^{\mathsf{T}} W_0 (P_N x). \end{aligned} \tag{15.17}$$

Under the linked law, the unused variables d_j, W_0 are independent of the complete scalar paths and the spins, because the link does not depend on them. Thus $\mathbb{E}[Z \mid \text{complete scalar paths and spins}] = 0$. The required integrability follows from the Gaussian moments of these variables and (15.7). Summing (15.16) and using Lemma 14.3 proves (15.11). □

Proposition 15.5 (Forward entropy at the fine horizon)**.** *Let $h = n^{-\mathsf{a}}$, $V = V_{\mathsf{f}}$ as in Lemma 14.3, with*

$$\mathsf{a} > \mathsf{f}/4, \quad \mathsf{a} + \mathsf{f}/2 < 1, \quad 0 < \mathsf{k}_{\mathsf{f}} < \min(\mathsf{a}, \mathsf{f}, 1 - \mathsf{a}). \tag{15.18}$$

For the primitive fine output Q_F,

$$\mathrm{KL}(Q_F \| \Pi_n) \le C_\beta n^{1-\mathsf{k}_{\mathsf{f}}}. \tag{15.19}$$

The disorder and spin marginals of Q_F are γ_n and Unif_n*, respectively.*

Proof. Combining the entropy–energy identity of Proposition 13.4, proved in [85, Lemma 4], with (15.6) and (15.11), we find

$$\begin{aligned} \mathrm{KL}(Q_F \| \Pi_n) \le C_\beta \Big[& nh + ne^{-V} + (N+1)(1+V) + \sqrt{N} \\ & + \frac{(N+1)^2 V R_V}{n} \Big]. \end{aligned} \tag{15.20}$$

Here we also used $\mathbb{E}_{\gamma_n} F^c \le n P_\beta$ and $p(V)^2 \le C_\beta e^{-V}$. By Lemmas 13.6 and 14.3, the powers of the five terms, up to fixed logarithmic factors, are $1 - \mathsf{a}$, $1 - \mathsf{f}$, a, $\mathsf{a}/2$, and $2\mathsf{a} + \mathsf{f}/2 - 1$. The inequality $\mathsf{a} + \mathsf{f}/2 < 1$ implies that the last power is smaller than a. Since the inequalities in (15.18) are strict, they absorb the constants and logarithmic factors. For the affine family of Section 19.5, these powers are $1/2, 1/2, 1/2, 1/4, 1/4$, and the claimed bound is $C_\beta n^{1/2+2\varepsilon_\omega}$. The energy estimate holds without restriction to a Gram event, including every outcome on which a query uses the fallback. The disorder marginal follows from averaging over the spins. The spin marginal is Unif_n because the linked scalar rows are independent and, by symmetry, $\mathbb{E} m_V = 0$ in each row. □

SECTION 16

Concentration on the growing mesh

In this section, we prove the concentration estimates used in Section 17. We use the population identities of Section 14.3 and the product law of the linked rows from Definition 14.2. Lemma 16.1 controls the two blocks of the empirical matrix. Lemma 16.6 controls the weighted errors that enter the energy expansion in Lemma 17.2.

Throughout, $V = V_{\mathsf{b}}$ and $N = N_{\mathsf{b}}$. All probability bounds are under the linked row law, averaged over the disorder. The equality of laws in Section 14.2 transports them to the source and replay experiments. Estimates involving only the paths are proved under the original law of the Brownian paths, which is also their linked marginal. Stochastic integrals use the spin-free filtration described there.

16.1 Errors and dyadic depth

We first define the errors used in both concentration estimates. For vectors in $\mathbb{R}^n$, we write $\langle u, v\rangle_n = n^{-1}\sum_{i=1}^n u_i v_i$. Let $\mathcal{T}_j$ and e_j be the query vectors of Section 14, and define

$$\begin{aligned} A_{\mathcal{T}} &= (\langle \mathcal{T}_i, \mathcal{T}_j\rangle_n), \quad q = (\langle \mathcal{T}_j, x\rangle_n), \quad v = (\langle e_j, x\rangle_n), \quad H = (\langle \mathcal{T}_i, e_j\rangle_n), \\ E_A &= A_{\mathcal{T}} - I, \qquad E_q = q - s, \qquad E_v = v - b, \qquad E_H = H - H_0. \end{aligned} \tag{16.1}$$

Here v is an empirical vector. We use $\|\cdot\|$ for the Euclidean norm of a vector and the operator norm of a matrix. Let $D_* = 1 + \lceil \log_2 N\rceil$ be the dyadic depth. We work in the case $N \geq 1$, which holds eventually; when there are no queries, the empty matrix formulas are interpreted accordingly.

Lemma 16.1 (Fixed Gram stability)**.** *For $n \geq 1$ and $1 \leq N < n$, let $R = R_V \geq 1$ and $D_* = 1 + \lceil \log_2 N\rceil$. For a linked row $Z = (\mathcal{T}, e, x)$, define $\widehat{\Sigma}_n = n^{-1}\sum_i Z_i Z_i^{\mathsf{T}}$ and $\Sigma = \mathbb{E} ZZ^{\mathsf{T}}$. Then*

$$\mathbb{P}\left\{\|E_A\| \vee \|E_H\| > (16 D_*^2)^{-1}\right\} \leq 2\,9^{2N+1} \exp\left\{-\frac{n}{2^{24} R D_*^4 \log(2+R)}\right\}. \tag{16.2}$$

The blocks E_A and E_H are coordinate compressions of $\widehat{\Sigma}_n - \Sigma$.

Proof. We bound the fourth moment of a row projection using its second moment and its sub-Gaussian tail. We then use a net to control the empirical covariance matrix,

following the empirical-Gram argument of [85, Section 4, proof of Proposition 5]. The variance-sensitive estimates below keep the dependence on R needed on the growing mesh.

Let $a = (a_{\mathcal{T}}, a_e, a_x)$ be a deterministic unit vector, and write $F = a^{\mathsf{T}} Z$. By the martingale representation, the normalized $\mathcal{T}$ projections have sub-Gaussian proxy R times the squared coefficient norm. The e projections are Gaussian, and x is a centered sign. By Hölder's inequality with exponents three, which allows dependence within a row, we have that F has proxy at most $3R$. By the population identities in Section 14.3 and $(u+v+w)^2 \le 3(u^2+v^2+w^2)$, we also have that $\mathbb{E}F^2 \le 3$.

Let $Y = F/\sqrt{3}$ and $Z' = F/\sqrt{3R}$. These have proxies R and 1, respectively, and $\mathbb{E}Y^2 \le 1$. We split $\mathbb{E}(YZ')^2$ at $(Z')^2 = Q = 16\log(2+R)$. The lower part is at most Q. For the upper part, we have

$$\mathbb{E}Y^4 \le 16R^2, \qquad \mathbb{E}\big[(Z')^4 \mathbf{1}_{\{(Z')^2 > Q\}}\big] \le 2e^{-Q/2}(Q^2+4Q+8).$$

By Cauchy–Schwarz, the upper part is at most $4\sqrt{2}R(Q+3)/(R+2)^4 < 2$. Hence $\mathbb{E}(YZ')^2 \le 20\log(2+R)$ and $\mathbb{E}F^4 \le 180R\log(2+R)$. Integrating the Gaussian tail, we also have $\|F^2\|_{\psi_1} \le 18R$, where the Orlicz norm is defined below. This implies that $X = F^2 - \mathbb{E}F^2$ has variance at most $M^2 = 180R\log(2+R)$ and Orlicz norm at most $S = 36R$.

Let $H = 4S\log(2+S)$. We split the lth moment at H. The lower part is at most $M^2 H^{l-2}$, and the upper part is at most

$$2H^l(2+S)^{-4} + 2l!(2S)^l(2+S)^{-2}.$$

For $M, S \ge 1$, the two upper terms are bounded by $8H^{l-2}$ and $8l!H^{l-2}$, respectively. Their sum with the lower part is at most $13l!M^2H^{l-2}$. Expanding the centered exponential series, we find

$$\log \mathbb{E}e^{\theta X} \le \frac{26M^2\theta^2}{2(1-H|\theta|)}.$$

Multiplying the moment generating functions of the rows, with $\theta = t/(26M^2+Ht)$ and both signs, we obtain the two-sided tail $2e^{-nt^2/[2(26M^2+Ht)]}$ for the empirical mean. The variance parameter is $4680R\log(2+R)$, and the scale is

$$144R\log(2+36R) \le 720R\log(2+R).$$

At tolerance $\varepsilon_0/2$, where $\varepsilon_0 = (16D_*^2)^{-1} \le 1$, the rate is at least $n\varepsilon_0^2/[40320R\log(2+R)]$.

A deterministic $1/4$ net of the unit sphere in dimension $2N+1$ has size at most 9^{2N+1}. The operator norm of a symmetric matrix is at most twice the largest absolute value of its quadratic form on this net [113, Corollary 4.2.13 and Exercise 4.4.3(b)]. Since $40320 \cdot 256 < 2^{24}$, we obtain (16.2) from the union bound over the net. Here we used that the rows have the product law of Definition 14.2, and that coordinate inclusions and projections have norm one. □

16.2 Weighted dyadic concentration

We next prove the weighted estimate. Four elementary estimates will control products, stochastic integrals with random brackets, weighted squares, and empirical means. We use them in Lemma 16.6, first for each dyadic node and then for the sum over the tree. The dyadic triangular decomposition is used in [85, Section 4, proof of Proposition 5]. Here the node estimates retain their durations before they are summed.

We write $\mathrm{SG}(s^2)$ for the centered bound $\mathbb{E}e^{\theta X} \le e^{s^2\theta^2/2}$ for every real θ, and let

$$\|X\|_{\psi_1} = \inf\{s > 0 : \mathbb{E}e^{|X|/s} \le 2\}.$$

The latter is a norm, so its triangle inequality holds also for dependent random variables; see [113, Section 2.7.1].

Lemma 16.2 (Products of sub-Gaussian variables)**.** *Let $R \ge 1$ and $u, s \ge 0$. If $\mathbb{E}Y^2 \le u^2$, Y is $\mathrm{SG}(Ru^2)$, and Z is $\mathrm{SG}(s^2)$, then*

$$\begin{aligned} &\mathbb{E}(YZ)^2 \le 20u^2s^2\log(2+R), \\ &\|YZ\|_{\psi_1} \le 10us\sqrt{R}, \qquad \|YZ - \mathbb{E}YZ\|_{\psi_1} \le 20us\sqrt{R}. \end{aligned} \tag{16.3}$$

The variables Y and Z may be dependent.

Proof. The sub-exponential product estimate is standard [113, Lemma 2.7.7]. We additionally retain the second moment of Y in the variance estimate. If $us = 0$, the assertions are immediate. Otherwise, we normalize to $u = s = 1$. By a union bound, we have that $\mathbb{P}(|YZ| > t) \le 4e^{-t/(2\sqrt{R})}$. Integrating this tail yields the two Orlicz bounds. To bound the second moment, we split at $Z^2 = Q = 16\log(2+R)$. The lower part is at most $Q\mathbb{E}Y^2$. For the upper part, we use Cauchy–Schwarz, and the Gaussian tails give

$$\mathbb{E}Y^4 \le 16R^2, \qquad \mathbb{E}[Z^4\mathbf{1}_{\{Z^2>Q\}}] \le 2e^{-Q/2}(Q^2+4Q+8).$$

The upper part is at most $4\sqrt{2}R(Q+3)/(R+2)^4 < 2$ for $R \ge 1$. In the last inequality, we used $Q + 3 \le 9(R+2)$ and $R/(R+2)^3 \le 1/27$. We note that the bound on the second moment uses both the second moment of Y and the two tails. This proves (16.3). □

Lemma 16.3 (A tail bound from the bracket)**.** *Let $\widetilde{M}$ be a continuous local martingale starting at zero, and suppose that $\|\langle\widetilde{M}\rangle_V\|_{\psi_1} \le A$. Then, for $t, u > 0$,*

$$\begin{aligned} \mathbb{P}(|\widetilde{M}_V| > t) &\le \mathbb{P}(\langle\widetilde{M}\rangle_V > u) + 2e^{-t^2/(2u)}, \\ \|\widetilde{M}_V\|_{\psi_1} &\le 5\sqrt{2A}. \end{aligned} \tag{16.4}$$

Proof. We stop at bracket level u and use the exponential supermartingale with parameters t/u and $-t/u$, the stopped stochastic-exponential argument of [64, Section 3.5]. On

$\{\langle \widetilde{M}\rangle_V \le u\}$, we have the first inequality. For the second, we take $u = t\sqrt{A/2}$ so that the resulting tail is bounded by $4e^{-t/\sqrt{2A}}$. Integrating it, we find that $\|\widetilde{M}_V\|_{\psi_1} \le 5\sqrt{2A}$. If $A = 0$, the martingale is zero and the claim is immediate. □

Lemma 16.4 (Weighted squares)**.** *Suppose that each Y_j is $\mathrm{SG}(s^2)$, where $s > 0$, and that $p_j \ge 0$ and $\sum_j p_j \le 1$. Then*

$$\mathbb{E}\exp\left\{\frac{\sum_j p_j Y_j^2}{6s^2}\right\} \le 2. \tag{16.5}$$

The variables Y_j may be dependent.

Proof. We use the sub-Gaussian-square estimate [113, Lemma 2.7.6] and Jensen's inequality, keeping the constants explicit. For a single Y_j, we integrate its Gaussian tail and bound the expectation by $1 + 2\int_1^\infty t^{-3}\,\mathrm{d}t = 2$. By Jensen's inequality, with the unused weight assigned to zero, we obtain (16.5). □

Lemma 16.5 (Variance-sensitive Bernstein bound)**.** *Suppose that $\mathbb{E}X = 0$, $\mathbb{E}X^2 \le M^2$, $\mathbb{P}(|X| > t) \le 2e^{-t/S}$, and $M, S \ge 1$. Let $H = 4S\log(2+S)$. Then, for integers $l \ge 2$,*

$$\mathbb{E}|X|^l \le 13\, l!\, M^2 H^{l-2}. \tag{16.6}$$

If $X_1, \dots, X_n$ are independent copies of X, then

$$\begin{aligned} \log \mathbb{E}e^{\theta X} &\le \frac{26M^2\theta^2}{2(1-H|\theta|)}, && |\theta| < H^{-1}, \\ \mathbb{P}\left(\left|\frac{1}{n}\sum_i X_i\right| > \eta\right) &\le 2\exp\left\{-\frac{n\eta^2}{2(26M^2 + H\eta)}\right\}, && \eta > 0. \end{aligned} \tag{16.7}$$

Proof. This is the Bernstein inequality in the form of [28, Theorem 2.10 and Corollary 2.11]; we include the calculation to record the constants needed here. In passing from the tail to the factorial moments, we keep the variance.

Splitting the lth moment at H and integrating the upper tail, we have

$$\mathbb{E}|X|^l \le M^2 H^{l-2} + 2H^l(2+S)^{-4} + 2l!(2S)^l(2+S)^{-2} \le 13\, l!\, M^2 H^{l-2}.$$

The boundary term is at most $8H^{l-2}$, and the last term is at most $8l!H^{l-2}$; here we used $H \ge 2S$ and $\log(2+S) \le (2+S)/2$. Expanding the exponential series, which is absolutely convergent, and using centering, we conclude that the first inequality in (16.7) holds. For the empirical mean, we use independence to multiply the moment generating functions. Taking $\theta = \eta/(26M^2 + H\eta)$, with both signs, gives the second inequality. □

Lemma 16.6 (Duration-normalized weighted concentration)**.** *Let $m_0 = 2^{\lceil \log_2 N\rceil}$ and $D_* = 1 + \log_2 m_0$. Pad the errors and s, b by zeros, and $A_{\mathcal{T}}$ by an identity, to dimension m_0. Dummy*

coordinates have duration zero. Let $\mathcal{D}$ be the complete dyadic tree on $\{1,\dots,m_0\}$, and let P_I denote coordinate projection onto I. Define

$$\begin{aligned} d_I &= \sum_{j\in I\cap[1,N]} h_{j-1}, \quad U_{\mathrm{dur}} = t_N, \\ \mathcal{N}_w(X)^2 &= \sum_{I\in\mathcal{D}} \|XP_Iw\|^2, \qquad L = \mathrm{Up}_{1/2}(E_A), \end{aligned} \tag{16.8}$$

where $\mathrm{Up}_{1/2}$ keeps the strict upper triangle and half its diagonal, and let

$$\begin{aligned} W_{\mathrm{wt}}^2 = \|E_q\|^2 + \|E_v\|^2 + \mathcal{N}_s(L)^2 + \mathcal{N}_s(L^{\mathsf{T}})^2 \\ + \mathcal{N}_b(L)^2 + \mathcal{N}_s(E_H)^2 + \mathcal{N}_s(E_H^{\mathsf{T}})^2. \end{aligned}$$

The duration d_I is that of the previous posterior cells; the query endpoint e_j has denominator $\sqrt{h_j}$ from the next cell. Let $R = R_V \ge 1$, $B_* = \max(1, \beta e^{1/8})$, and

$$\begin{aligned} &\ell_R = \log(2+R), \quad M_*^2 = 25B_*^2D_*^2\ell_R, \\ &S_* = 20B_*D_*\sqrt{R}, \qquad \mathsf{v} = 26M_*^2, \qquad H_* = 4S_*\log(2+S_*). \end{aligned} \tag{16.9}$$

For every $n, N \ge 1$, every deterministic grid with used cells at most one and $h_j/h_{j-1} \le e^{1/4}$, and every $\eta > 0$, under the linked row product law,

$$\mathbb{P}\left\{W_{\mathrm{wt}} > 2\eta\sqrt{2+5D_*U_{\mathrm{dur}}}\right\} \le 2(10m_0-3)5^{m_0}\exp\left\{-\frac{n\eta^2}{2(\mathsf{v}+H_*\eta)}\right\}. \tag{16.10}$$

With this failure bound, the following estimates hold simultaneously:

$$\begin{aligned} &\|E_q\|, \|E_v\| \le 2\eta, \\ &\|LP_Is\|, \|L^{\mathsf{T}}P_Is\|, \|LP_Ib\|, \|E_HP_Is\|, \|E_H^{\mathsf{T}}P_Is\| \le 2\eta\sqrt{d_I} \quad (I\in\mathcal{D}). \end{aligned} \tag{16.11}$$

Zero-duration nodes have zero error and are omitted when dividing by $\sqrt{d_I}$. On the growing grid with $h = n^{-\mathsf{a}}$, $V = V_{\mathsf{b}}$, and $\mathsf{a} > \mathsf{b}/4$, there exists a fixed finite C such that

$$\mathbb{P}\{W_{\mathrm{wt}} > C_\beta n^{-\mathsf{w}}\log^C n\} \le 2\exp\left\{C_\beta n^{\mathsf{a}} - \frac{c_\beta n^{1-\max\{2\mathsf{w},\mathsf{w}+\mathsf{b}/4\}}}{\log^C n}\right\}, \tag{16.12}$$

$$\mathbb{P}\{\|E_A\| \vee \|E_H\| > (16D_*^2)^{-1}\} \le 2\exp\left\{C_\beta n^{\mathsf{a}} - \frac{c_\beta n^{1-\mathsf{b}/2}}{\log^C n}\right\}. \tag{16.13}$$

Constants are independent of deterministic row directions, tree nodes, the selected cut, and the rotation angle. The finite row estimate allows all N; its transport to the n-spin proposal uses $N < n$, which holds eventually on the parameter family since $\mathsf{a} < 1$.

Proof. We first estimate projections of a single row. We decompose each triangular projection by dyadic depth and separate its half diagonal. After dividing each interval projection by $\sqrt{d_I}$, we use Lemma 16.5 and a sphere net. Summing the resulting squared bounds over the tree gives (16.10).

Row projections and the durations. For a deterministic α, the population identities and martingale representation in Section 14.3 imply that

$$\begin{aligned} &\mathbb{E}(\alpha^{\mathsf{T}}\mathcal{T})^2 = \|\alpha\|^2, \\ &\qquad \alpha^{\mathsf{T}}\mathcal{T} \text{ is } \mathrm{SG}(R\|\alpha\|^2), \qquad \alpha^{\mathsf{T}}e \text{ is } \mathrm{SG}(\|\alpha\|^2). \end{aligned} \tag{16.14}$$

Indeed, $\alpha^{\mathsf{T}}\mathcal{T}$ is the integral with integrand $\alpha_j\chi_t/s_j$ on previous cell j; its bracket is bounded by $R\|\alpha\|^2$. For $u_j = w_j s_j$ with $|w_j| \le B_*$, we have

$$Z_J = \sum_{j\in J} u_j\mathcal{T}_j = \sum_{j\in J} w_j \int_{t_{j-1}}^{t_j} \chi_t \,\mathrm{d}M_t \text{ is } \mathrm{SG}(B_*^2 d_I), \qquad J \subset I. \tag{16.15}$$

The bracket bound is deterministic. This implies that the same estimate holds conditionally before its supporting future cells. Further, $\|P_I s\|^2 \le d_I$, so $s^{\mathsf{T}}P_I e$ is $\mathrm{SG}(d_I)$. The weights b_j satisfy

$$b_j \le \beta\sqrt{h_j r(t_j)}, \qquad s_j \ge \sqrt{h_{j-1} r(t_j)}, \qquad \frac{b_j}{s_j} \le \beta\sqrt{\frac{h_j}{h_{j-1}}} \le \beta e^{1/8}. \tag{16.16}$$

Here we used the monotonicity of r and $p \le 1$. Note that the previous and next durations enter this calculation separately.

The strict triangular part. Fix a deterministic unit direction α in the padded dimension and one node I. At a dyadic depth, let (A, B) range over the pairs of disjoint left and right children of its parents. The projections of the strict upper triangle at this level and of its transpose are, respectively,

$$\begin{aligned} N_+ &= \sum_{(A,B)} Y_A Z_{B\cap I}, \quad Y_A = \sum_{j\in A} \alpha_j \mathcal{T}_j, \\ N_- &= \sum_{(A,B)} Z_{A\cap I} Y_B, \quad Y_B = \sum_{j\in B} \alpha_j \mathcal{T}_j. \end{aligned}$$

The first is the predictable integral with multiplier $Y_A w_j \chi_t$ on future cells $j \in B \cap I$. The second has multiplier $Z_{A\cap I}\alpha_j\chi_t/s_j$ on future cells $j \in B$. The future supports at one depth are disjoint. By conditional centering before each future support, the block products are orthogonal, although their past and future factors need not be independent. The isometry for N_+ and Lemma 16.2 for each product in N_- give

$$\mathbb{E}N_+^2 \le B_*^2 d_I \sum_A \|P_A\alpha\|^2 \le B_*^2 d_I, \qquad \mathbb{E}N_-^2 \le 20 B_*^2 d_I \ell_R. \tag{16.17}$$

Their brackets satisfy

$$\langle N_+\rangle \le B_*^2 d_I \sum_A Y_A^2, \qquad \langle N_-\rangle \le R \sum_B \|P_B\alpha\|^2 Z_{A\cap I}^2. \tag{16.18}$$

Using Lemma 16.4 with weights $\|P_A\alpha\|^2$ and $\|P_B\alpha\|^2$, respectively, we bound the Orlicz norms of both brackets by $6B_*^2 R d_I$. Lemma 16.3 then implies that the Orlicz norms of both projections are at most $18B_*\sqrt{Rd_I}$. There are $D_* - 1$ depths, and we sum the bounds over them. By the triangle inequalities in L^2 and in the Orlicz norm, this costs at most a factor equal to the number of depths.

The half diagonal. In either orientation, the half diagonal is

$$Q_I = \frac{1}{2}\sum_{j\in I} \frac{\alpha_j w_j}{s_j}((\Delta m_j)^2 - s_j^2) = N_I + D_I, \tag{16.19}$$

where

$$N_I = \sum_{j\in I} \frac{\alpha_j w_j}{s_j} \int_{t_{j-1}}^{t_j} (m_t - m_{t_{j-1}})\chi_t \,\mathrm{d}M_t,$$
$$D_I = \frac{1}{2}\sum_{j\in I} \frac{\alpha_j w_j}{s_j}\left(\int_{t_{j-1}}^{t_j} \chi_t^2 \,\mathrm{d}t - s_j^2\right), \qquad \Delta m_j = m_{t_j} - m_{t_{j-1}}.$$

Both terms have mean zero, and D_I is a centered random drift. By isometry,

$$\mathbb{E}N_I^2 \le B_*^2 \sum_{j\in I} \alpha_j^2 h_{j-1} \le B_*^2 d_I.$$

Let $S_j = \sup_{[t_{j-1},t_j]} |m_t - m_{t_{j-1}}|$. By the maximal inequality for the exponential martingale, we have $\mathbb{P}(S_j > u) \le 2e^{-u^2/(2h_{j-1})}$, so $\|S_j^2\|_{\psi_1} \le 6h_{j-1}$. The bracket of N_I is bounded by $B_*^2 R\sum \alpha_j^2 S_j^2$, whose Orlicz norm is at most $6B_*^2 R d_I$. By Lemma 16.3, we have $\|N_I\|_{\psi_1} \le 18B_*\sqrt{Rd_I}$. For $q_j = \int_{t_{j-1}}^{t_j} \chi_t^2 \,\mathrm{d}t$, we have

$$\|q_j - s_j^2\|_2 \le \|q_j\|_2 \le \int_{t_{j-1}}^{t_j} \sqrt{r(t)}\,\mathrm{d}t \le \sqrt{h_{j-1}}s_j,$$

where we used $\chi^4 \le \chi^2$. It follows that

$$\begin{aligned} \|D_I\|_2 &\le \tfrac{1}{2}B_*\sqrt{d_I}, \qquad |D_I| \le \tfrac{1}{2}B_*\sqrt{Rd_I}, \\ \mathbb{E}Q_I^2 &\le \tfrac{9}{4}B_*^2 d_I, \qquad \|Q_I\|_{\psi_1} \le 20B_*\sqrt{Rd_I}. \end{aligned} \tag{16.20}$$

These estimates keep track of the unequal durations of the cells. In particular, the variance bound involves d_I without a power of R, which would be lost if each duration were replaced by the maximal one.

The remaining projections and normalization. The two projections of the full cross blocks are the centered products $(\alpha^\mathsf{T}\mathcal{T})(s^\mathsf{T}P_I e)$ and $(\alpha^\mathsf{T}e)(s^\mathsf{T}P_I\mathcal{T})$. By (16.14)–(16.16) and Lemma 16.2, their variances are at most $20d_I\ell_R$, and their Orlicz norms are at most

$20\sqrt{Rd_I}$. Here the centering subtracts the block H_0. The two projections without intervals, $\alpha^{\mathsf{T}}(\mathcal{T}x - s)$ and $\alpha^{\mathsf{T}}(ex - b)$, have variance at most one and Orlicz norm at most $10\sqrt{R}$, since multiplication by the terminal sign preserves squares and absolute values.

Combining (16.17)–(16.20), we find that each of the five interval projections, after division by $\sqrt{d_I}$, and each of the two projections without intervals is centered and satisfies

$$\mathbb{E}X^2 \le M_*^2, \qquad \|X\|_{\psi_1} \le S_*. \tag{16.21}$$

For a single row, the five vectors indexed by an interval are

$$\mathrm{Up}_{1/2}(\mathcal{T}\mathcal{T}^{\mathsf{T}} - I)P_I s, \qquad \mathrm{Up}_{1/2}(\mathcal{T}\mathcal{T}^{\mathsf{T}} - I)^{\mathsf{T}}P_I s, \qquad \mathrm{Up}_{1/2}(\mathcal{T}\mathcal{T}^{\mathsf{T}} - I)P_I b,$$
$$(\mathcal{T}e^{\mathsf{T}} - H_0)P_I s, \qquad (e\mathcal{T}^{\mathsf{T}} - H_0^{\mathsf{T}})P_I s.$$

To pad these vectors, we extend their errors by zero on the dummy coordinates. The identity matrix in their row formulas acts on the first N coordinates.

We have $M_*, S_* \ge 1$ in (16.21), even for arbitrarily small d_I. Lemma 16.5 therefore applies after the division by $\sqrt{d_I}$, although its hypotheses need not hold before this normalization.

Nets and the growing grid. Fix a deterministic $1/2$ net of the unit sphere of size at most 5^{m_0}. For every vector z, we have $\|z\| \le 2\max|\alpha^{\mathsf{T}}z|$, where the maximum is over this net [113, Corollary 4.2.13 and Exercise 4.4.2]. There are $2 + 5(2m_0 - 1) = 10m_0 - 3$ vectors to control. Using (16.7) and the union bound, we conclude that (16.11) holds. At every depth, the durations partition t_N, so $\sum_I d_I = D_* U_{\mathrm{dur}}$. Summing the squared errors before normalization, we obtain

$$W_{\mathrm{wt}}^2 \le 4\eta^2(2 + 5D_* U_{\mathrm{dur}}),$$

which proves (16.10). The union bound is over $(10m_0 - 3)5^{m_0}$ events, and the nets are in dimension m_0.

For the stopped growing grid, Lemma 14.3 implies that $N \le 8n^{\mathsf{a}}$, $D_* = O(\log n)$, and $U_{\mathrm{dur}} \le V = O(\log n)$. By Lemma 13.6, we have $R \le C_\beta n^{\mathsf{b}/2}(\log n)^{1/2}$. Then M_*^2 is bounded by a fixed power of $\log n$, and $H_* \le C_\beta n^{\mathsf{b}/4}\log^C n$. Taking $\eta = n^{-\mathsf{w}}$ in (16.10) and using $x/(u+v) \ge (1/2)\min(x/u, x/v)$ yields (16.12). To prove (16.13), we use Lemma 16.1 for the compressions E_A, E_H. Its rate is n/R divided by a fixed power of $\log n$, which gives the separate loss $\mathsf{b}/2$. By the equality of row laws in Section 14.2, all these bounds also hold under the linked source and replay laws. □

SECTION 17

The joint cancellation and coarse domination

The goal of this section is to prove Proposition 17.8, which bounds the restricted coarse output kernel by an unnormalized Gibbs weight. We use the Gram and weighted concentration estimates of Lemmas 16.1 and 16.6. The deterministic energy expansion and the joint information–energy identity proved below give the accuracy needed for this bound. The resulting kernel estimate is used in Section 18 to control the connecting sequence. This sharpens the likelihood comparison of [85, Section 4, Proposition 5, equation (6)], used for the coarse proposals in [85, Section 6, proof of Proposition 7].

Unless stated otherwise, $V = V_{\mathrm{b}}$ and $N = N_{\mathrm{b}} \geq 1$. Throughout, we use the source and replay laws of Definition 14.2, and the empirical Gram quantities $A_{\mathcal{T}}, q, v, H, E_A, E_q, E_v, E_H$ of Section 16. Here v is an empirical vector, distinct from the scalar time. We recall that H_0 is strictly lower triangular, $\|s\|, \|b\|, \|H_0\| \leq 1$, and $e_h = s^{\mathsf{T}} b$. Matrix norms are operator norms, and norms of coefficient vectors in $\mathbb{R}^N$ are Euclidean. For the query vectors in $\mathbb{R}^n$, we write $\langle u, v \rangle_n = n^{-1} u^{\mathsf{T}} v$.

17.1 Dyadic algebra and the row telescope

We first record the deterministic estimates used in the energy expansion. Let $m_0 = 2^{\lceil \log_2 N \rceil}$ and $D_* = 1 + \log_2 m_0$. As in Lemma 16.6, errors and coefficient vectors are padded by zeros to dimension m_0, and the Gram matrix is padded by an identity block. For the complete dyadic tree $\mathcal{D}$ and its coordinate projections P_I, we use

$$\mathcal{N}_w(X)^2 = \sum_{I \in \mathcal{D}} \|X P_I w\|^2, \qquad L = \mathrm{Up}_{1/2}(E_A),$$

where $\mathrm{Up}_{1/2}$ keeps the strict upper triangle and half of the diagonal, and Up keeps the full upper triangle. The weighted error W_{wt} is defined in Lemma 16.6. In this section L denotes this triangular matrix. The dyadic decomposition of triangular truncations and the analysis of the Cholesky factor by half-diagonal truncation already appear in the proof of [85, Section 4, Proposition 5]. Here we retain the weighted projections to obtain a quadratic remainder.

Lemma 17.1 (Deterministic dyadic estimates)**.** *For matrices X, Y and vectors a, b, w of the appropriate dimensions,*

$$\begin{aligned} |a^{\mathsf{T}} \operatorname{Up}_{1/2}(X^{\mathsf{T}}Y)b|,\ |a^{\mathsf{T}} \operatorname{Up}(X^{\mathsf{T}}Y)b| &\le 2\mathcal{N}_a(X)\mathcal{N}_b(Y), \\ \mathcal{N}_w(\operatorname{Up}_{1/2}(X^{\mathsf{T}}Y)) &\le 2D_*\|X\|\mathcal{N}_w(Y), \\ \mathcal{N}_w(\operatorname{Up}_{1/2}(X^{\mathsf{T}}Y)^{\mathsf{T}}) &\le 2D_*\|Y\|\mathcal{N}_w(X), \\ \mathcal{N}_w(X) &\le \sqrt{D_*}\|X\|\|w\|, \\ \|\operatorname{Up}_{1/2}(X)\|,\ \|\operatorname{Up}(X)\| &\le D_*\|X\|. \end{aligned} \tag{17.1}$$

Suppose $\|E_A\| \vee \|E_H\| \le \varepsilon_0 = (16D_*^2)^{-1}$. Let C be the upper triangular positive-diagonal Cholesky factor of $I + E_A$, and define $U = C - I$, $B = C^{-\mathsf{T}}$, and $c = B(s + E_q)$. Then

$$U = L - \operatorname{Up}_{1/2}(U^{\mathsf{T}}U), \tag{17.2}$$

$$\|U\| \le 2D_*\varepsilon_0, \qquad \|B\| \le 2, \qquad 2D_*\|U\| \le 1/2, \qquad D_*\|E_H\| \le 1. \tag{17.3}$$

Further, $\mathcal{N}_w(U) \le 2\mathcal{N}_w(L)$ for $w = s, b$, $\|U^{\mathsf{T}}s\| \le 2W_{\mathrm{wt}}$, and, with $u = c - s$,

$$\begin{aligned} u &= B(E_q - U^{\mathsf{T}}s), \qquad \|u\| \le 6W_{\mathrm{wt}}, \\ u - E_q + L^{\mathsf{T}}s &= \operatorname{Up}_{1/2}(U^{\mathsf{T}}U)^{\mathsf{T}}s - U^{\mathsf{T}}u. \end{aligned} \tag{17.4}$$

Proof. We decompose the strict upper triangle into dyadic rectangles, and then use the Cholesky equation to control the nonlinear term of the Cholesky factor. Each rectangle in the scalar estimates contributes $\langle XP_{I_{\mathrm{left}}}a, YP_{I_{\mathrm{right}}}b\rangle$. Cauchy–Schwarz over rectangles and then over singletons gives the factor two in (17.1).

For the vector estimate, fix an input interval I. The external output $P_{<I}X^{\mathsf{T}}YP_Iw$ is orthogonal to the internal output and has squared norm at most $\|X\|^2\|YP_Iw\|^2$. At each internal depth, the outputs of the left children have disjoint supports. Cauchy–Schwarz over depths implies that the squared norm of the internal output is at most $D_*\|X\|^2$ times the sum of the squared norms of the right descendants and the singletons. When we sum over I, each descendant is counted at most D_* times. The resulting factor $1 + 2D_*^2$ in squared norm is at most $(2D_*)^2$. Reversing left and right proves the transpose estimate. Partitioning at each depth gives the generic bound for $\mathcal{N}_w(X)$. Since the rectangle supports at one depth are disjoint, we also have the truncation bound.

The Cholesky equation $E_A = U + U^{\mathsf{T}} + U^{\mathsf{T}}U$ implies (17.2). Along $I + tE_A$, $0 \le t \le 1$, the Cholesky factor is continuous, as follows from the finite recursion that defines it by positive square roots. Writing $x(t)$ for the norm of this factor minus I, we have

$$x(t) \le D_*\varepsilon_0 + D_*x(t)^2, \qquad x(0) = 0.$$

At $x = 2D_*\varepsilon_0$, the right side is strictly smaller than x, since $4D_*^2\varepsilon_0 = 1/4$. By continuity, we have (17.3). Applying the first vector estimate in (17.1) to (17.2) and absorbing the

nonlinear term, we obtain $\mathcal{N}_w(U) \le 2\mathcal{N}_w(L)$ for $w = s, b$. The transpose estimate at the root implies that $\|U^{\mathsf{T}} s\| \le 2W_{\mathrm{wt}}$. Since $C^{\mathsf{T}} c = s + E_q$, we obtain both identities in (17.4), and $\|u\| \le 2(W_{\mathrm{wt}} + 2W_{\mathrm{wt}}) = 6W_{\mathrm{wt}}$. □

Lemma 17.2 (Weighted energy and discarded Gaussian variances)**.** *If* $\|E_A\| \vee \|E_H\| \le (16D_*^2)^{-1}$*, the queried energy is*

$$\begin{aligned} E_{\mathrm{qry}} &= c^{\mathsf{T}}(b + E_v) - c^{\mathsf{T}} \operatorname{Up}\big(B(H_0 + E_H)\big)c, \\ C^{\mathsf{T}}C &= I + E_A, \qquad B = C^{-\mathsf{T}}, \qquad c = B(s + E_q), \end{aligned} \tag{17.5}$$

where C is upper triangular with positive diagonal. Define

$$\ell(E) = E_q^{\mathsf{T}} b - s^{\mathsf{T}} \operatorname{Up}_{1/2}(E_A) b + s^{\mathsf{T}} E_v - s^{\mathsf{T}} \operatorname{Up}(E_H) s. \tag{17.6}$$

Then

$$\|c - s\| \le 6W_{\mathrm{wt}}, \qquad |E_{\mathrm{qry}} - e_h - \ell(E)| \le 200 D_* W_{\mathrm{wt}}^2. \tag{17.7}$$

Conditionally on the complete scalar paths and signs, the two unused energy terms in the source reconstruction are independent centered Gaussians with variances

$$\frac{1}{2n}\sum_j c_j^4, \qquad \frac{(1 - \|c\|^2)^2}{2n}. \tag{17.8}$$

If $\|c - s\| \le \tau$ and $t_N \ge v_*$, their sum Z satisfies

$$\operatorname{Var}(Z \mid \textit{paths, signs}) \le \frac{C_\beta}{n}\{h + \tau^2 + p(t_N)^2\}. \tag{17.9}$$

These conditional assertions are taken before imposing any event that inspects the unused Gaussian variables.

Proof. We first expand the queried energy using Lemma 17.1, and then compute the two conditional Gaussian variances. Since the target Gram matrix is positive definite, no query uses the deterministic fallback. We write $\mathcal{T}^{\mathrm{col}} = (\mathcal{T}_1, \dots, \mathcal{T}_N)$ and $\mathcal{U} = (u_1, \dots, u_N)$ for the $n \times N$ matrices with these columns, so that $n^{-1}\mathcal{U}^{\mathsf{T}}\mathcal{U} = I_N$. The Gram–Schmidt factorization is $\mathcal{T}^{\mathrm{col}} = \mathcal{U}C$, with $C^{\mathsf{T}}C = A_{\mathcal{T}}$. Then $\mathcal{U} = \mathcal{T}^{\mathrm{col}}C^{-1}$ and $c = Bq$. The (i, j) entry of BH is $\langle u_i, e_j\rangle_n$. Subtracting the first j query projections of e_j in the energy gives (17.5), with truncation after multiplication. Since B is lower triangular and H_0 is strictly lower triangular,

$$\operatorname{Up}(BH_0) = 0. \tag{17.10}$$

Indeed, a nonzero summand in an entry of the upper triangle would require $k \le i \le j < k$. Thus the population block cancels. If we pad the errors, H_0, and all vectors by zeros, and the Gram matrix by an identity block, then by uniqueness the Cholesky factor becomes $C \oplus I$. The padding preserves the norms, products, truncations, and energy.

Let $u = c - s$ as in Lemma 17.1. The remainder of $c^{\mathsf{T}}(b + E_v)$ after its linear terms is

$$s^{\mathsf{T}} \operatorname{Up}_{1/2}(U^{\mathsf{T}}U)b - (Ub)^{\mathsf{T}}u + u^{\mathsf{T}}E_v.$$

By (17.1), its absolute value is at most $26W_{\mathrm{wt}}^2$. For the E_H term, we first replace c by s on both sides. The three resulting errors are bounded by

$$2\sqrt{D_*}\|u\|\mathcal{N}_s(E_H), \qquad 2\sqrt{D_*}\|u\|\mathcal{N}_s(E_H^{\mathsf{T}}), \qquad D_*\|E_H\|\|u\|^2.$$

Their sum is at most $(24\sqrt{D_*} + 36)W_{\mathrm{wt}}^2$. Finally, let $F_0 = B - I$. Since $F_0^{\mathsf{T}} = -B^{\mathsf{T}}U$ and $\|F_0\| \le 2\|U\|$, Lemma 17.1 implies that

$$\mathcal{N}_c(F_0^{\mathsf{T}}) \le 7W_{\mathrm{wt}}, \qquad \mathcal{N}_c(E_H) \le 7W_{\mathrm{wt}}.$$

By the scalar estimate, we have $|c^{\mathsf{T}} \operatorname{Up}(F_0E_H)c| \le 98W_{\mathrm{wt}}^2$. Adding the bounds, we find that the total is

$$(160 + 24\sqrt{D_*})W_{\mathrm{wt}}^2 \le 200D_*W_{\mathrm{wt}}^2,$$

which proves (17.7).

In the source reconstruction of Section 14.1, the variables $d_j \sim N(0, 2/n)$ and W_0 are independent of all complete scalar paths and terminal signs. Neither the scalar update nor the terminal link reads them. The two terms of Z in the energy decomposition of Lemma 15.4 have the conditional variances in (17.8). On the scalar event in the statement, we have $\|c\| \le 1$ by orthonormality, and

$$\sum_j c_j^4 \le \max_j c_j^2 \le 2\max_j s_j^2 + 2\tau^2 \le C_\beta h + 2\tau^2.$$

Here we used that $s_j^2 \le h_{j-1}r(t_{j-1}) \le C_\beta h$, which follows from the exponential tail envelope and the growing durations. Also, $\|s\|^2 = \int_0^{t_N} r = 1 - p(t_N)$ for $t_N \ge v_*$, so

$$1 - \|c\|^2 \le p(t_N) + 2\tau.$$

Adding (17.8) proves (17.9). Since the unused primitives are independent of the complete scalar paths, these laws hold conditionally on the paths and terminal signs. □

The linear statistic (17.6) can be written as an average over the scalar rows.

Lemma 17.3 (Row telescope)**.** *Let $A_t = xm_t - m_t^2/2$ and $w_j = b_j/s_j$, and define*

$$\zeta_h = \sum_{j=1}^{N} w_j(A_{t_j} - A_{t_{j-1}}) + \sum_{j=1}^{N} s_j(x - m_{t_j})e_j - \frac{1}{2}e_h. \tag{17.11}$$

Then

$$e_h + \ell(E) = \frac{1}{n}\sum_{i=1}^{n} \zeta_{h,i}, \qquad \mathbb{E}_{\nu_V}\zeta_h = e_h. \tag{17.12}$$

Proof. Expanding the triangles and using $\sum_{l\le j} s_l \mathcal{T}_l = m_{t_j}$ gives the first identity. To check the constant, we note that the empirical expression is the row average of

$$\begin{aligned} xb^{\mathsf{T}}\mathcal{T} + xs^{\mathsf{T}}e - s^{\mathsf{T}}\,\mathrm{Up}_{1/2}(\mathcal{T}\mathcal{T}^{\mathsf{T}})b \\ - s^{\mathsf{T}}\,\mathrm{Up}(\mathcal{T}e^{\mathsf{T}})s - e_h/2. \end{aligned}$$

The first and third terms combine into $\sum_j w_j \Delta A_j$, and the other pair gives the second sum in (17.11). The population expectation of $\mathrm{Up}_{1/2}(\mathcal{T}\mathcal{T}^{\mathsf{T}})$ is $I/2$, while that of $\mathrm{Up}(\mathcal{T}e^{\mathsf{T}})$ is zero. Since $\mathbb{E}x\mathcal{T} = s$ and $\mathbb{E}xe = b$, the expectation is $e_h + e_h - e_h/2 - e_h/2 = e_h$. This verifies (17.12), including the term $-e_h/2$. □

17.2 The joint information–energy cancellation

We next compare the row telescope with the continuous scalar energy. The Itô identities below and the information integral of Lemma 15.1 imply an identity at the terminal time. We estimate its terminal fluctuation and its mesh residual separately. The scalar entropy–energy balance is [85, Lemma 4], and the auxiliary-reference likelihood is from [85, Section 4, proof of Proposition 5]. The joint terminal identity below compares the fluctuations of these quantities on the same row history. Its terminal correction and the growing-mesh residual are derived here.

Lemma 17.4 (Itô formula for the time-dependent field)**.** *On every finite horizon, under the original scalar Wiener law,*

$$\mathrm{d}U(t, Y_t) = m_t\,\mathrm{d}M_t + \frac{1}{2}\bar{\alpha}(t)m_t^2\,\mathrm{d}t, \qquad \mathrm{d}m_t = \chi_t\,\mathrm{d}M_t. \tag{17.13}$$

The identities hold across CDF atoms and the support endpoint. They continue to hold after adjoining an independent initial sign, and their fixed-horizon almost-sure identities transfer to the linked row law.

Proof. By Proposition 4.3 and the scalar estimates of Section 5, the Parisi solution used in Proposition 13.4 is continuous in time with its spatial derivatives, smooth in space on compact rectangles, and absolutely continuous in time there. Its time derivative is $-(U_{yy} + \bar{\alpha}U_y^2)/2$ almost everywhere, and its spatial derivative satisfies the differentiated equation.

To justify Itô's formula directly, we stop Y on leaving $[-A, A]$, extend the time fields slightly at the endpoints, and mollify in time. The fields and spatial derivatives converge uniformly on compact subrectangles. The time derivatives converge at Lebesgue-almost-every time, uniformly on compact spatial sets: the CDF has only countably many discontinuities, and the other factors are continuous. The derivatives are bounded on the

stopped rectangle. The classical Itô formula [64, Theorem 3.3.6] holds for the mollifications, and we pass to the limit using dominated convergence for the drift integrals and the isometry for the stochastic integrals. The drift for U is $\bar{\alpha}m^2/2$, and the drift for m is $-(m_{yy} + 2\bar{\alpha}m\chi)/2 + \bar{\alpha}m\chi + m_{yy}/2 = 0$. Passing to an almost-sure subsequence and then using rational times and continuous versions, we obtain the identities at all times. As $A \to \infty$, every continuous path on a finite horizon is eventually unstopped. Since $m, \chi, \bar{\alpha}$ are bounded, the integrands displayed above satisfy the required integrability conditions. There is no integral against $\mathrm{d}\bar{\alpha}$, so the atoms produce no jump term. By uniqueness, the second identity agrees with Lemma 13.5. Since the terminal link is positive and bounded, these identities hold under the linked law, with the integrals still defined in the original Brownian filtration. □

Lemma 17.5 (Joint information–energy terminal cancellation)**.** *For a deterministic finite $V \geq v_*$, define*

$$\begin{aligned} A_t &= xm_t - m_t^2/2, \\ Z_V &= \beta \int_0^V p(t)\,\mathrm{d}A_t + \int_0^V \sqrt{r(t)}(x - m_t)\,\mathrm{d}B_t, \\ \zeta_\infty &= Z_V - e(V)/2. \end{aligned} \tag{17.14}$$

All stochastic integrals here and in the following lemma are first defined under the unlinked Brownian pair law with an independent initial uniform sign, using the measurable left-sum versions of Section 15.1. Then, under ν_V,

$$I_V - \beta\zeta_\infty = \log 2 - P_\beta - \frac{\beta^2 p(V)^2}{4} + \frac{\beta^2 p(V)}{2}(x - m_V)^2. \tag{17.15}$$

The centered terminal fluctuation

$$T_V = \frac{\beta^2 p(V)}{2}\{(x - m_V)^2 - p(V)\}$$

satisfies $|T_V| \leq 2\beta^2 p(V)$ and

$$\begin{aligned} \mathrm{Var}_{\nu_V}(T_V) &= \frac{\beta^4 p(V)^2}{4}\{4p(V) - 3r(V) - p(V)^2\} \\ &\leq \beta^4 p(V)^3. \end{aligned} \tag{17.16}$$

The identity (17.15) *holds at the endpoint $V \geq v_*$.*

Proof. We first match the information and energy at the support endpoint, and then extend the identity beyond support and compute the terminal variance. Before support, we have $\rho = 1$ and $\sigma = 0$, so the query driver B equals M. By (17.13), $\beta^2 p' = -\bar{\alpha}$, and $A_0 = Y_0 = 0$, we have

$$\mathrm{d}\{\beta^2 pA + xY - U(t, Y)\} = \beta^2 p\,\mathrm{d}A + (x - m)\,\mathrm{d}B = \beta\,\mathrm{d}Z.$$

Indeed, the extra drift is $\bar{\alpha}(-A + xm - m^2/2) = 0$. Integrating up to v_* yields

$$\beta Z_{v_*} = \beta^2 p_* A_* + x Y_* - U(v_*, Y_*) + U(0,0).$$

At the support endpoint,

$$m_* = \tanh Y_*, \qquad \beta^2 p_* = \beta^2 - v_*, \qquad U(v_*, y) = \log(2\cosh y) + (\beta^2 - v_*)/2.$$

Since $I_{v_*} = \log(1 + x m_*)$, $\log(1 + x \tanh y) = xy - \log\cosh y$, and $A + (x-m)^2/2 = 1/2$, we obtain

$$I_{v_*} - \beta Z_{v_*} = \log 2 - U(0,0) + \frac{1}{2}\beta^2 p_* (x - m_*)^2. \tag{17.17}$$

After support, $\chi = 1 - m^2$ and $\xi = x - m$. By the information integral of Lemma 15.1, we have

$$\begin{aligned} \mathrm{d}(I - \beta Z) &= -\beta^2 p \,\mathrm{d}A - \frac{1}{2}\beta^2 r (x-m)^2 \,\mathrm{d}t \\ &= \mathrm{d}\{\tfrac{1}{2}\beta^2 p (x-m)^2\}, \end{aligned}$$

where we used $p' = -r$. This extends (17.17) to every $V \geq v_*$.

With the time change of Proposition 13.4, the normalization of the pressure in (4.5) reads

$$P_\beta = U(0,0) - \frac{1}{2\beta^2}\int_0^{\beta^2} t\bar{\alpha}(t)\,\mathrm{d}t.$$

This implies the identity

$$\beta e(V) + \frac{1}{2}\beta^2 p(V)^2 = 2\{U(0,0) - P_\beta\}, \qquad V \geq v_*. \tag{17.18}$$

To check it, note that the left side is constant after support. At support it equals $\int_0^{v_*} p + \beta^2 p_*^2/2$. Using integration by parts and $p' = -\bar{\alpha}/\beta^2$, we find that

$$\int_0^{v_*} p = v_* p_* + \beta^{-2}\int_0^{v_*} t\bar{\alpha}(t)\,\mathrm{d}t.$$

The remaining expression is

$$v_* p_* + \beta^2 p_*^2/2 = \frac{\beta^4 - v_*^2}{2\beta^2} = \beta^{-2}\int_{v_*}^{\beta^2} t\,\mathrm{d}t,$$

which proves (17.18). Substituting $\zeta_\infty = Z_V - e(V)/2$ proves (17.15).

Given the terminal path, the sign has mean m_V. Averaging over this sign, we find that

$$\begin{aligned} \mathbb{E}[(x - m_V)^2 \mid \text{path}] &= 1 - m_V^2, \\ \mathbb{E}[(x - m_V)^4 \mid \text{path}] &= (1 - m_V^2)(1 + 3m_V^2). \end{aligned}$$

For $V \geq v_*$, their expectations are $p(V)$ and $4p(V) - 3r(V)$. This proves (17.16), including the centering. At a fixed finite such V, the variance is positive: the conditional sign variance of $(x - m_V)^2$ is $4m_V^2(1 - m_V^2)$, which is positive with positive probability since the scalar diffusion is nondegenerate. Thus the centered row as a whole cannot have variance $O(h)$ uniformly at fixed V as $h \to 0$.

The restriction to the terminal endpoint is necessary. If the identity extended to times before support, then $I_t - \beta Z_t - \beta^2 p(t)(x - m_t)^2/2$ would be constant. The coefficient of $\mathrm{d}B$ in its differential is $\xi_t - (x - m_t)$. At time zero, $\chi_0 = p(0) = 1/\beta$, which implies that this coefficient is $x(1/\beta - 1) \neq 0$ and has positive bracket on a short interval. It is the matching at v_* in (17.17) that makes the terminal identity possible. □

Lemma 17.6 (Growing-mesh residual MGF)**.** *On the growing mesh of Section 14.1, suppose $0 < h \leq 1$, all used cells are at most one, and $V = t_{N+1} \geq v_*$. Define*

$$R_h = Z_V - (\zeta_h + e_h/2).$$

There exist $C_\beta, c_\beta > 0$, independent of h, N, V, such that

$$\begin{aligned} \log \mathbb{E}_{\nu_V} e^{\theta(R_h - \mathbb{E}R_h)} &\leq C_\beta \theta^2 h, \qquad |\theta| \leq c_\beta/h, \\ \mathbb{E}R_h &= \frac{3}{2}(e(V) - e_h). \end{aligned} \tag{17.19}$$

For every $n \geq 1$ and $0 < \eta \leq 1$, independent identically distributed linked rows obey

$$\mathbb{P}\left\{ \left| \frac{1}{n} \sum_i (R_{h,i} - \mathbb{E}R_h) \right| > \eta \right\} \leq 2e^{-c_\beta n \eta^2/h}. \tag{17.20}$$

Proof. We decompose the residual into four terms, estimate their centered MGFs, and then compute the mean of the residual. Both omitted endpoint cells are retained in the decomposition.

The residual decomposition. On previous cells $[t_{j-1}, t_j)$, $1 \leq j \leq N$, let $C(t) = \beta p(t) - w_j$; on the final cell let $C(t) = \beta p(t)$. On current cells $[t_j, t_{j+1})$, $j \geq 1$, let $G(t) = \sqrt{r(t)} - s_j/\sqrt{h_j}$; on the auxiliary cell zero, let $G(t) = \sqrt{r(t)}$. We write $m_{\lfloor t \rfloor} = m_{t_j}$ on cell j, including $m_0 = 0$, and define

$$\begin{aligned} \mathcal{M} &= \int_0^V C\chi \,\mathrm{d}M + \int_0^V G \,\mathrm{d}B, \\ \mathcal{N} &= -\int_0^V Cm\chi \,\mathrm{d}M - \int_0^V Gm_{\lfloor t \rfloor} \,\mathrm{d}B, \\ \mathcal{Q} &= -\frac{1}{2}\int_0^V C\chi^2 \,\mathrm{d}t, \qquad \mathscr{L}_j = \int_{t_j}^{t_{j+1}} \sqrt{r(t)}(m_t - m_{t_j}) \,\mathrm{d}B_t. \end{aligned}$$

Then

$$R_h = x\mathcal{M} + \mathcal{N} + \mathcal{Q} - \sum_{j=0}^{N} \mathscr{L}_j. \tag{17.21}$$

This follows from $\mathrm{d}A = (x-m)\chi\,\mathrm{d}M - \chi^2\,\mathrm{d}t/2$ and

$$\begin{aligned}\sqrt{r}(x-m_t) - \frac{s_j}{\sqrt{h_j}}(x-m_{t_j}) \\ = G(t)(x-m_{t_j}) - \sqrt{r}(m_t - m_{t_j}).\end{aligned}$$

The initial Brownian cell and the final posterior increment, which is not queried, are two distinct omissions, and both occur in (17.21). The current cell of the last query remains included in e_h.

Uniform bounds on the growing mesh. By the continuation kernel of Lemma 13.6, we have the envelope $p(t), r(t) \le C_\beta e^{-t/2}$ for all $t \ge 0$, after increasing C_β. In particular, $\int_0^\infty e^{t/4}(p+r)\,\mathrm{d}t < \infty$. For this mesh, we have

$$\begin{gathered}(N+1)h \le 8, \qquad \max_j h_j\sqrt{r(t_j)} \le \sqrt{C_\beta}h, \\ \sum_j h_j^2 r(t_j) \le 8C_\beta h.\end{gathered} \tag{17.22}$$

Indeed, $e^{-t_j/4} - e^{-t_{j+1}/4} \ge h/8$ for $h_j \le 1$, and $h_j^2 e^{-t_j/2} = h^2$. These two facts also bound the number of queries.

The logarithmic profile bounds of Proposition 13.4, with $h_j/h_{j-1} = e^{h_{j-1}/4}$, imply, for t in previous cell j and u in current cell j, respectively,

$$\begin{aligned}\left|\log\frac{w_j}{\beta p(t)}\right| &\le (33/8)h_j, \\ -h_{j-1}/8 \le \log\frac{s_j}{\sqrt{h_j r(u)}} &\le (15/8)h_j.\end{aligned} \tag{17.23}$$

For the first bound, we compare the integrand $p\sqrt{r}$ throughout the current cell to its value on the previous cell and use $\sqrt{h_j/h_{j-1}} = e^{h_{j-1}/8}$. Each logarithm varies by at most the elapsed time. For the second, we compare the average of r on the previous cell to $r(u)$ and use the same duration ratio. These comparisons imply that

$$\begin{aligned}|C(t)| &\le C_\beta h_j p(t), \qquad t \in [t_{j-1}, t_j), \\ |G(t)| &\le C_\beta h_j\sqrt{r(t)}, \quad t \in [t_j, t_{j+1}),\ j \ge 1.\end{aligned} \tag{17.24}$$

Indeed, the inequality $|e^x - 1| \le e^L|x|$ for $|x| \le L$ proves this with $L = 33/8$ and $15/8$. On a previous cell, $h_j \le e^{1/4} h e^{t/4}$, and on a current cell, $h_j \le h e^{t/4}$. These bounds imply

$$\int C^2 \, \mathrm{d}t \le C_\beta h, \qquad \int G^2 \, \mathrm{d}t \le C_\beta h, \qquad |\mathcal{Q}| \le C_\beta h. \tag{17.25}$$

In detail, the integral of C^2 over the interior cells is bounded by $C_\beta h^2 \int e^{t/2} p(t)^2 \, \mathrm{d}t = O_\beta(h^2)$, and the final cell contributes at most $C_\beta h_N e^{-t_N} \le C_\beta h$. For G^2, the auxiliary cell zero contributes $O_\beta(h)$; elsewhere we use $h_j^2 \le h_j \le h e^{t/4}$ and integrate $e^{t/4} r$. To bound $\mathcal{Q}$, we integrate $|C|$, using $h \int e^{t/4} p$ and the bound $h_N p(t_N) \le C_\beta h$ on the final cell. All constants are uniform in the horizon. Since $|\chi|, |m|, |m_{\lfloor t \rfloor}| \le 1$ and $|\rho| \le 1$, both path martingales $\mathcal{M}, \mathcal{N}$ have deterministic bracket at most $C_\beta h$.

The lag integrals. For cell j of length h_j, we condition only on the past $\mathcal{F}^{\mathrm{path}}_{t_j}$ of the original spin-free Brownian pair. The increment $m_t - m_{t_j}$ has bracket at most h_j. We stop it at its first crossing of $\sqrt{u h_j}$. The outer lag integral, stopped at this time, has bracket at most $u h_j^2 r(t_j)$, since r is nonincreasing. By the stopped exponential-martingale argument [64, Section 3.5] and a union bound, we have, for $u > 0$,

$$\begin{aligned} \mathbb{P}\{ |\mathscr{L}_j| > u h_j \sqrt{r(t_j)} \mid \mathcal{F}^{\mathrm{path}}_{t_j} \} &\le 4 e^{-u/2}, \\ \mathbb{E}[\mathscr{L}_j \mid \mathcal{F}^{\mathrm{path}}_{t_j}] &= 0. \end{aligned} \tag{17.26}$$

Integrating this tail and expanding the centered series, we obtain

$$\mathbb{E}[e^{\theta \mathscr{L}_j} \mid \mathcal{F}^{\mathrm{path}}_{t_j}] \le e^{32 \theta^2 h_j^2 r(t_j)}, \qquad |\theta| \le [4 h_j \sqrt{r(t_j)}]^{-1}. \tag{17.27}$$

Indeed, for every integer $l \ge 2$, the conditional lth absolute moment is at most $4 l! [2 h_j \sqrt{r(t_j)}]^l$. Beyond the linear term, which vanishes, the series of the conditional MGF is at most $16 \theta^2 h_j^2 r(t_j) / (1 - 2|\theta| h_j \sqrt{r(t_j)})$. Iterating over cells and using (17.22), we find that the centered log MGF is bounded by $C_\beta \theta^2 h$ on $|\theta| \le c_\beta / h$ for $\sum_j \mathscr{L}_j$. The iteration uses only the conditional bounds and needs no independence between cells. The isometry also gives the variance identity $\mathbb{E}\mathscr{L}_j^2 = (1/2)(\int_{t_j}^{t_{j+1}} r)^2$. The tail and MGF estimates above are conditional on the past of the original pair; this variance identity is unconditional.

The centered MGF and the mean. Linking does not change the law of the paths. In particular, $\mathcal{N}$ has its usual sub-Gaussian bound under ν_V. Multiplying $\mathcal{M}$ by x preserves its absolute tail, but the product need not be a martingale. We use the standard centering principle for sub-Gaussian variables [113, Lemma 2.6.8], verifying an MGF constant by symmetrization. To obtain a centered MGF without a prefactor, let X' be an independent

copy of $X = x\mathcal{M}$. By Jensen's inequality, $\mathbb{E}e^{\theta(X-\mathbb{E}X)} \le \mathbb{E}e^{\theta(X-X')}$. If X has tail $2e^{-u^2/(2v)}$, the symmetric difference has tail $4e^{-u^2/(8v)}$ and even moments at most $4(8v)^l l!$. Since $(2l)! \ge (l!)^2$, its series is at most $4e^{8v\theta^2} - 3 \le e^{32v\theta^2}$. This applies with $v = C_\beta h$. By Hoeffding's lemma [28, Lemma 2.2], we have a centered log MGF bounded by $C_\beta\theta^2h^2$ for $\mathcal{Q}$, using (17.25). We combine these bounds and (17.27) in (17.21) by Hölder's inequality with four factors. Since the number of factors is fixed, this proves the MGF bound in (17.19).

For the mean, we have, on every deterministic grid, the identities

$$\int_0^V Cr = e(V) - e_h, \qquad \int_0^V G\rho p = e(V) - e_h. \tag{17.28}$$

The first uses the previous-cell value s_j^2, and the second the current-cell value $\sqrt{h_j}b_j$. By Lemma 13.5 and posterior averaging, $\mathbb{E}[x\mathcal{M}] = \int Cr + \int G\rho p$, $\mathbb{E}\mathcal{N} = \mathbb{E}\mathscr{L}_j = 0$, and $\mathbb{E}\mathcal{Q} = -(1/2)\int Cr$. We conclude that $\mathbb{E}R_h = 3(e(V) - e_h)/2$. The same integrals of the coefficients show that $|e(V) - e_h| \le C_\beta h$, in agreement with Lemma 14.3. For independent identically distributed rows, the centered MGFs multiply. The exponential Markov inequality with $\theta = c\eta/h$, for a sufficiently small fixed $c > 0$, proves (17.20) for all $0 < \eta \le 1$.

Integral versions and replay. At fixed finite V and fixed query cells, we refine within each cell by dyadic left sums, retaining its endpoints. For the lag sums, the isometry and $\mathbb{E}(m_t - m_u)^2 \le t - u$ bound the squared error by $8V\omega_{\sqrt{r}}(\varepsilon')^2 + V\varepsilon' \to 0$, where $\omega_{\sqrt{r}}$ is the modulus of continuity of $\sqrt{r}$ on this fixed horizon. The integrals with deterministic weights and the likelihood integral against D are identified with limits of left sums in the same way. Passing to an almost-sure subsequence and using Fatou's lemma, we obtain the same MGF bound for the limit as for the finite sums. With bounded test functions of the past, we obtain the conditional version of Fatou's lemma in the same way.

The query driver is $B = \int \rho\,\mathrm{d}M + \int \sigma\,\mathrm{d}\widetilde{M}$, with the independent orthogonal driver of Section 13.4. We retain both terms. Freezing ρ, σ at the endpoints of the refinement produces a squared error at most $4\int[(\rho_{\varepsilon'} - \rho)^2 + (\sigma_{\varepsilon'} - \sigma)^2]\,\mathrm{d}t \to 0$. Before support, we use $\mathrm{d}m = \chi\,\mathrm{d}M$. We do not split a cell that crosses the support endpoint, since splitting it and restarting its lag baseline would omit a term. The identity of the full histories in Definition 14.2 and the uniqueness of these sums identify the replay versions, including the reference of Lemma 15.1. For fixed n, the linking transfers null sets. The MGF and tail estimates follow from the unchanged path law and sign symmetrization, without a factor 2^n. The conditioning in (17.26) and (17.27) never involves the terminal sign, the final matrix, the futures of the other rows, or any crossing event. □

17.3 Coarse domination

We combine the terminal cancellation and the residual estimate to bound the information minus the queried energy. The weighted remainder and the unused Gaussian energy then give the bound on the restricted kernel in Proposition 17.8.

Lemma 17.7 (Concentration of the row $I - \beta\zeta_h$)**.** *For $V \geq v_*$, we have the exact identity*

$$\begin{aligned} I_V - \beta\zeta_h = \log 2 - P_\beta - \frac{\beta^2 p(V)^2}{4} + \frac{\beta^2 p(V)}{2}(x - m_V)^2 \\ + \beta R_h + \frac{\beta}{2}(e_h - e(V)). \end{aligned} \tag{17.29}$$

Its mean is

$$j(V) - \beta e_h = \log 2 - P_\beta + \frac{\beta^2 p(V)^2}{4} + \beta(e(V) - e_h), \tag{17.30}$$

and its centered random part is $T_V + \beta(R_h - \mathbb{E}R_h)$, with variance at most $C_\beta\{p(V)^3 + h\}$. In particular, for $V = V_{\mathsf{b}}$, $h = n^{-\mathsf{a}}$, $\mathsf{z} < \min(\mathsf{a}, \mathsf{b}/2)$, and all sufficiently large n,

$$\mathbb{P}\left\{\frac{1}{n}\sum_i (I_{V,i} - \beta\zeta_{h,i}) > \log 2 - P_\beta + C_\beta n^{-\mathsf{z}}\right\} \leq 2e^{-c_\beta n^{1+\mathsf{a}-2\mathsf{z}}}. \tag{17.31}$$

Proof. By (17.14) and the definition of R_h, we have $\zeta_h = \zeta_\infty - R_h + (e(V) - e_h)/2$. Inserting this into (17.15) yields (17.29). Using (17.19) and $\mathbb{E}(x - m_V)^2 = p(V)$, we obtain (17.30). This also agrees with the information mean of Lemma 15.1 and with (17.12). Combining the centered decomposition with (17.16) and (17.19), we conclude that the variance has the stated bound; this does not require the two terms to be independent.

The terminal correction is deterministically at most $2\beta^2 p(V) = O_\beta(n^{-\mathsf{b}/2}) = o(n^{-\mathsf{z}})$. Its mean and the deterministic mesh error $O_\beta(h)$ fit within the same accuracy. By (17.20) with $\eta = n^{-\mathsf{z}}$, absorbing the fixed coefficient β into its constants, we obtain (17.31). The terminal variance remains $O(p(V)^3)$. Bernstein's inequality for bounded variables [28, Theorem 2.10 and Corollary 2.11] would give it the rate $n\min(\eta^2/p(V)^3, \eta/p(V))$; here we use only its deterministic amplitude. □

Proposition 17.8 (Intrinsic coarse domination)**.** *Suppose the fine conditions of Proposition 15.5 hold and, in addition,*

$$\begin{gathered} \mathsf{a} > \mathsf{b}/4, \qquad 0 < \mathsf{k}_{\mathsf{b}} < \mathsf{z} < \min(\mathsf{a}, \mathsf{b}/2, 2\mathsf{w}), \\ \max\{2\mathsf{w}, \mathsf{w} + \mathsf{b}/4, \mathsf{b}/2, 2\mathsf{z} - \mathsf{a}, 2\mathsf{z} - 2\mathsf{w}, 2\mathsf{z} - \mathsf{b}\} < \mathsf{s}_{\mathsf{b}}, \\ \mathsf{a} < 1 - \mathsf{s}_{\mathsf{b}}. \end{gathered} \tag{17.32}$$

Let μ_n^{aux} be the independent product law of the spin-free replay inputs $(D, B^0, \mathrm{br}, (C_j)_j)$ from Definition 14.2, and let U be independent and uniform on $[0,1]^n$. Only the inputs through V_{b} are

read here; unused coordinates extending the experiment to the fine horizon may be retained. We write $m_i^{g,a}(V_{\mathsf{b}})$ *for the resulting replay terminal magnetizations and define*

$$X_B(g,a,U)_i = \mathbf{21}\left\{U_i \le \frac{1+m_i^{g,a}(V_{\mathsf{b}})}{2}\right\} - 1.$$

There exists a Borel replay event E_B *such that, eventually,*

$$\mathbb{P}_{\gamma_n \otimes \mu_n^{\mathrm{aux}} \otimes \mathrm{d}U}(E_B^c) \le e^{-n^{1-\mathsf{s_b}}}. \tag{17.33}$$

The event includes $\|A_{\mathcal{T}} - I\| \le 3/4$*. Its restricted output kernel satisfies, for* γ_n*-almost every* g*, simultaneously for all* $x \in \Sigma_n$*,*

$$\begin{aligned} q_B^{\mathrm{good}}(g,x) &= \int \mathbf{1}_{E_B}\mathbf{1}_{\{X_B(g,a,U)=x\}}\,\mathrm{d}\mu_n^{\mathrm{aux}}(a)\,\mathrm{d}U \\ &\le e^{-nP_\beta + n^{1-\mathsf{k_b}}} e^{\beta H_g^c(x)}. \end{aligned} \tag{17.34}$$

The event E_B *contains no pressure event, and* (17.33) *is an annealed failure bound.*

Proof. We intersect the weighted and Gram events with bounds on the unused Gaussian energy and the row residual. The row identity gives a likelihood bound on this event, which we then disintegrate over the disorder.

The energy errors. At $V = V_{\mathsf{b}}$, we take the simultaneous weighted event of Lemma 16.6 at $\eta = n^{-\mathsf{w}}$ and the fixed Gram event $\|E_A\| \vee \|E_H\| \le (16D_*^2)^{-1}$. On their intersection, by Lemmas 16.6 and 17.2, we have

$$\begin{aligned} \|c - s\| &\le \tau_n = C_\beta n^{-\mathsf{w}} \log^C n, \\ |E_{\mathrm{qry}} - e_h - \ell(E)| &\le C_\beta n^{-2\mathsf{w}} \log^C n. \end{aligned} \tag{17.35}$$

The failure probabilities of the two events have the respective raw rates $n^{1-\max\{2\mathsf{w},\mathsf{w}+\mathsf{b}/4\}}/\operatorname{polylog} n$ and $n^{1-\mathsf{b}/2}/\operatorname{polylog} n$, less the cost $C_\beta n^{\mathsf{a}}$ of the net. By the strict inequalities in (17.32), these rates eventually dominate both the cost of the net and any fixed multiple of $n^{1-\mathsf{s_b}}$.

For the discarded energy, we use its conditional law before restricting it by its own tail. The endpoint of the last previous cell satisfies $t_N = \mathsf{b}\log n + o(1)$, since $V_{\mathsf{b}} - t_N = h_N = o(1)$. Thus $p(t_N)^2 \le C_\beta n^{-\mathsf{b}}$. On the event in (17.35), we see from (17.9) that its conditional variance is at most

$$C_\beta n^{-1}\{n^{-\mathsf{a}} + n^{-2\mathsf{w}}\log^C n + n^{-\mathsf{b}}\}.$$

The conditional probability that $|Z| > n^{-\mathsf{z}}$ is at most

$$2\exp\left\{-\frac{c_\beta n^{1-\max\{2\mathsf{z}-\mathsf{a},\,2\mathsf{z}-2\mathsf{w},\,2\mathsf{z}-\mathsf{b}\}}}{\log^C n}\right\}. \tag{17.36}$$

If the conditional variance is zero, this probability is zero. Integrating over the event on paths and signs in (17.35) gives the same bound for the probability of its intersection with this failure. Here we used the sum of the two unused Gaussian terms, with the conditional variance computed above. The coarser bound $1/(2n)$ would not give the required rate at these exponents.

The replay event. We intersect the preceding events with $|Z| \le n^{-\mathsf{z}}$ and

$$\left| \frac{1}{n} \sum_i (R_{h,i} - \mathbb{E} R_h) \right| \le n^{-\mathsf{z}}.$$

The latter fails with probability at most $2e^{-c_\beta n^{1+\mathsf{a}-2\mathsf{z}}}$ by (17.20), with loss $2\mathsf{z} - \mathsf{a}$. All these events are replay events: the Gram blocks and scalar histories are replay functions, the residual uses the integral versions of Lemma 15.1, and $Z = H_g^c(x)/n - E_{\mathrm{qry}}$. We use the Borel versions fixed by the conventions of Section 13. The identity between the source and replay laws in Definition 14.2 transports their probabilities. The sum of the finitely many failure probabilities is eventually at most $e^{-n^{1-\mathsf{s}_\mathsf{b}}}$, because each raw exponent exceeds $1 - \mathsf{s}_\mathsf{b}$ and the exponent of the net by a strict margin. We denote the resulting event by E_B.

On E_B, (17.12), (17.29), and the mean in (17.19) imply that

$$\begin{aligned} \frac{1}{n} I^{(n)} - \log 2 - \frac{\beta}{n} H_g^c(x) &\le -P_\beta + C_\beta \{ p(V_\mathsf{b}) + p(V_\mathsf{b})^2 + h \\ &\qquad + n^{-\mathsf{z}} + n^{-2\mathsf{w}} \log^C n \} \\ &\le -P_\beta + n^{-\mathsf{k}_\mathsf{b}} \end{aligned}$$

eventually. Here the terminal correction has amplitude $O_\beta(p(V_\mathsf{b})) = O_\beta(n^{-\mathsf{b}/2})$, the deterministic mean of R_h is $O_\beta(h)$, and the weighted remainder is controlled by (17.35). Since $\mathsf{k}_\mathsf{b} < \mathsf{z} < \min(\mathsf{a}, \mathsf{b}/2, 2\mathsf{w})$, all constants and logarithms are absorbed. The terminal variance (17.16) is controlled here by its deterministic amplitude.

Disintegration. On this replay space, $\Lambda = e^{I^{(n)}} R_D$ and $(g, x)_* R_D = \gamma_n \otimes \mathrm{Unif}_n$. For every Borel set A of primitive matrices and each spin,

$$\Lambda(E_B,\ g \in A,\ x) \le \int_A e^{-nP_\beta + n^{1-\mathsf{k}_\mathsf{b}} + \beta H_g^c(x)}\, \gamma_n(\mathrm{d}g).$$

The factor 2^n from the logarithmic inequality cancels the factor 2^{-n} of the reference law of the spin. Realizing the linked spin by terminal thresholds does not change the restricted law. Disintegration over the disorder, whose marginal is unchanged, gives (17.34). Since the spin space is finite, there exists one common measurable null set for all spins. We redefine the restricted kernel to be zero there. This does not affect its later use,

since the nondegenerate Gaussian regressions used later are absolutely continuous with respect to γ_n. This proves the bound on the subprobability kernel and, separately, the annealed failure estimate. In the coupled experiment of Section 18.3, the auxiliary inputs are independent, and this will justify applying the proposition after the outer target is adjoined. □

SECTION 18

Smoothing, coverage, and the connecting sequence

The goal of this section is to prove Proposition 18.9, which excludes balanced cuts with small stationary flow on the event $\|W(g)\|_{\mathrm{op}} \le 12$. We use the fine entropy estimate of Proposition 15.5 to obtain coverage of both sides of a cut, and the coarse domination estimate of Proposition 17.8 to construct a sequence that avoids its enlarged boundary. These are the smoothing and connecting sequence arguments of [85, Sections 5–6, Propositions 6–7], with the estimates at the fine and coarse horizons established in Sections 15 and 17. In Section 19, the witness estimate and Lemma 13.2 will imply Theorem 1.3.

Throughout, we fix $\beta > 1$ and the exponent parameters before taking $n \to \infty$, as in Section 13.3. We retain the primitive disorder, completed Hamiltonian, and joint law Π_n of Section 13.1, and the source and replay constructions of Section 14.

18.1 Smoothing and coverage

We first record the transport inequality for disorder conditioned on a spin. Under Π_n, the spin marginal is Unif_n: this follows from the gauge transformation $g_{ij} \mapsto x_i x_j g_{ij}$ and summation over spins. Conditional on x, primitive disorder has density

$$p_x(g) = 2^n (2\pi)^{-n^2/2} \exp\{-\|g\|_F^2/2 + \beta H_g^c(x) - F^c(g)\}. \tag{18.1}$$

Since F^c is a log-sum-exp of linear functions and $H_g^c(x)$ is linear in g, we have that the negative logarithm of p_x is 1-strongly convex. We write W_2 for the quadratic Wasserstein distance for the Euclidean norm on the primitive matrix space, that is, the Frobenius norm.

Lemma 18.1. *For every $x \in \Sigma_n$ and every probability measure η on the primitive matrix space,*

$$W_2(\eta, p_x)^2 \le 2\,\mathrm{KL}(\eta \| p_x). \tag{18.2}$$

Proof. This is the transport inequality for measures with a strongly convex potential; see [24, Corollary 3.10] and its use in [85, Section 5]. We recall the finite-dimensional argument. For a smooth positive density η, let T be the optimal transport map from p_x to η, which is the gradient of a convex function by the convex-gradient transport theorem

of Brenier [32] in the full Euclidean generality of [73, Main Theorem], and let V_x be the negative logarithm of p_x. By the change-of-variables formula, we have

$$\mathrm{KL}(\eta \| p_x) = \int \big(V_x(T(g)) - V_x(g) - \log\det DT(g)\big)\, p_x(g)\, \mathrm{d}g.$$

By strong convexity and $-\log\det DT \geq \mathrm{tr}(I - DT)$, we have

$$\mathrm{KL}(\eta \| p_x) \geq \frac{1}{2} \int \|T(g) - g\|_F^2 p_x(g)\, \mathrm{d}g.$$

In the last step, we used integration by parts to cancel the linear terms. For a general measure with finite entropy, we obtain the inequality by smooth truncation and regularization, followed by lower semicontinuity. If the entropy is infinite, there is nothing to prove. □

For every L-Lipschitz linear function f and every $t \in \mathbb{R}$, the variational formula for relative entropy [28, Corollary 4.14] and Lemma 18.1 imply that

$$\mathbb{E}_{p_x} \exp\{t(f - \mathbb{E}_{p_x} f)\} \leq e^{t^2 L^2/2}. \tag{18.3}$$

Indeed, $\mathbb{E}_\eta f - \mathbb{E}_{p_x} f \leq L W_2(\eta, p_x)$, and maximizing $|t| L\sqrt{2K} - K$ over $K \geq 0$ proves (18.3).

Let $\delta = n^{-\mathsf{D}}$, and write the fine output as $Q_F(\mathrm{d}g, x) = \gamma_n(\mathrm{d}g) q_F(g, x)$. We take the version of q_F given by the replay integral, with the conventions of Section 13.1, and define

$$\nu_g^{\mathrm{sm}}(x) = \int q_F(\sqrt{1-\delta}\, g + \sqrt{\delta}\, z, x)\, \gamma_n(\mathrm{d}z). \tag{18.4}$$

For every g, we have that ν_g^{sm} is a probability measure on the spins.

Lemma 18.2. *Let $g_\star, g_\circ$ be independent with law γ_n, and set $g_{\mathrm{base}} = \sqrt{1-\delta}\, g_\star + \sqrt{\delta}\, g_\circ$. Then*

$$\mathbb{E}_{g_\star, g_\circ} \mathrm{KL}(\pi_{g_\star} \| \nu_{g_{\mathrm{base}}}^{\mathrm{sm}}) \leq \frac{\beta^2 (n-1)}{2} \delta + \delta^{-1} \mathrm{KL}(Q_F \| \Pi_n). \tag{18.5}$$

In particular, if $\mathsf{u}_{\mathrm{E}} < \min(\mathsf{D}, \mathsf{k}_{\mathrm{f}} - \mathsf{D})$, the outer event

$$\mathcal{C}_n = \big\{(g_\star, g_\circ) : \mathrm{KL}(\pi_{g_\star} \| \nu_{g_{\mathrm{base}}}^{\mathrm{sm}}) \leq n^{1-\mathsf{u}_{\mathrm{E}}}\big\} \tag{18.6}$$

has complementary probability at most $C_\beta(n^{\mathsf{u}_{\mathrm{E}} - \mathsf{D}} + n^{\mathsf{u}_{\mathrm{E}} - \mathsf{k}_{\mathrm{f}} + \mathsf{D}}) = o(1)$.

Proof. We first compare the joint laws after smoothing, conditioning on the spin. We then compare the Gibbs laws at the target and its smoothing base. This is the reverse comparison of [85, Section 5, Proposition 6]; we use Gaussian integration by parts for the second comparison.

We disintegrate Π_n and Q_F over the spin, whose marginal is uniform under both laws. For each spin, by (18.2), we can couple the conditional laws of the disorder with squared cost at most twice their forward relative entropy. Under the channel $g \mapsto \sqrt{1-\delta}\, g + \sqrt{\delta}\, Z$, the relative entropy between the Gaussian output laws at two inputs is $(1-\delta)\|g - g'\|_F^2/(2\delta)$ in either direction. By convexity of relative entropy, applied to the coupled mixture and averaged over the common spin, we have

$$\mathrm{KL}(\Pi_n^{\mathrm{sm}} \| Q_F^{\mathrm{sm}}) \le \delta^{-1} \mathrm{KL}(Q_F \| \Pi_n). \tag{18.7}$$

This comparison for Gaussian convolutions also appears in [70, Lemma 11]. Both smoothed measures have disorder marginal γ_n. By Gaussian reversibility, we identify the conditional spin law of Q_F^{sm} with (18.4), and we denote that of Π_n^{sm} by $\bar{\pi}_g$. Under the joint law of $(g_\star, g_{\mathrm{base}})$, expanding the logarithms of the probabilities on the finite spin space gives the entropy chain rule [28, Chapter 4], in the form used in [85, Section 5, proof of Proposition 6]:

$$\begin{aligned} \mathbb{E}\,\mathrm{KL}(\pi_{g_\star} \| \nu_{g_{\mathrm{base}}}^{\mathrm{sm}}) = \mathbb{E}\,\mathrm{KL}(\pi_{g_\star} \| \bar{\pi}_{g_{\mathrm{base}}}) \\ + \mathbb{E}\,\mathrm{KL}(\bar{\pi}_{g_{\mathrm{base}}} \| \nu_{g_{\mathrm{base}}}^{\mathrm{sm}}). \end{aligned} \tag{18.8}$$

Using the same decomposition with any alternative kernel at g_{base}, we find that the first term is at most $\mathbb{E}\,\mathrm{KL}(\pi_{g_\star} \| \pi_{g_{\mathrm{base}}})$. When intermediate expressions may be infinite, we justify these identities by finite approximations, or by the variational formula for relative entropy and monotone approximation. By Proposition 15.5 and the following calculation, the upper bounds are finite.

Since $g_\star$ and g_{base} have the same marginal law, the expected difference of log partition functions vanishes. Using the independence of $g_\circ$, we obtain

$$\begin{aligned} \mathbb{E}\,\mathrm{KL}(\pi_{g_\star} \| \pi_{g_{\mathrm{base}}}) &= \beta(1 - \sqrt{1-\delta})\, \mathbb{E}_{g_\star} \sum_x \pi_{g_\star}(x) H_{g_\star}^c(x) \\ &\le \tfrac{1}{2}\beta^2 (n-1)(1 - \sqrt{1-\delta}) \le \tfrac{1}{2}\beta^2 (n-1)\delta. \end{aligned} \tag{18.9}$$

For the middle inequality, Gaussian integration by parts gives the off-diagonal contribution $(\beta/n)\sum_{i<j}(1 - \langle x_i x_j \rangle_{g_\star}^2)$; the spin-independent diagonal has mean zero. Since the derivatives that occur here are bounded spin moments and Gaussian first moments are finite, the integrations are justified. Combining (18.7)–(18.9) proves (18.5). The final assertion follows from Proposition 15.5 and Markov's inequality. □

Lemma 18.3. *On $\mathcal{C}_n$, every spin has positive $\nu_{g_{\mathrm{base}}}^{\mathrm{sm}}$ probability, and, simultaneously for every $S \subseteq \Sigma_n$,*

$$\pi_{g_\star}(S) \ge \epsilon_n = n^{-\mathsf{d}} \quad \Longrightarrow \quad \nu_{g_{\mathrm{base}}}^{\mathrm{sm}}(S) \ge \exp\{-n^{\mathsf{d}}(n^{1-\mathsf{u}_{\mathrm{E}}} + \log 2)\}. \tag{18.10}$$

If $\mathsf{e} + \mathsf{d} < \mathsf{u}_{\mathrm{E}}$, *two independent draws from* $\nu^{\mathrm{sm}}_{g_{\mathrm{base}}}$ *cross every* ϵ_n*-balanced cut with probability at least*

$$\varpi_n = \exp\{-2n^{\mathsf{d}}(n^{1-\mathsf{u}_{\mathrm{E}}} + \log 2)\} \geq e^{-2n^{1-\mathsf{e}}} \tag{18.11}$$

eventually.

Proof. This is the two-point entropy-contraction argument of [85, Section 5, Proposition 6]; the partition bound is [28, Theorem 4.20]. Finite relative entropy and the strict positivity of $\pi_{g_\star}(x)$ imply that $\nu^{\mathrm{sm}}_{g_{\mathrm{base}}}(x) > 0$ for every spin. We pass to the partition $\{S, S^c\}$, which does not increase relative entropy. If $a = \pi_{g_\star}(S)$ and $b = \nu^{\mathrm{sm}}_{g_{\mathrm{base}}}(S)$, the binary relative entropy is at least $a \log(1/b) - \log 2$, since binary entropy is at most $\log 2$. We have proved (18.10). Applying it to both sides of a balanced cut and multiplying the independent probabilities gives (18.11); the strict power gap also absorbs the $\log 2$ term. The implication holds for every set once the outer pair is fixed, including a cut selected from this pair. □

18.2 The witness event and the enlarged boundary

We now select a balanced cut with small flow, whenever one exists, and bound the Gibbs mass of its enlarged boundary. The selection and the comparison of boundary weights follow [85, Section 6, Proposition 7]. Here the enlargement cost is controlled by concentration of sums of local fields over small sets of sites under the conditional disorder law. We derive this sparse-field estimate below from Lemma 18.1 and its linear-function MGF consequence; the transport input is [24, Corollary 3.10].

For $\epsilon > 0$ and $\phi > 0$, we define the witness event

$$\begin{aligned}\mathcal{W}_n(\epsilon, \phi) = \{g : \|W(g)\|_{\mathrm{op}} \leq 12,\\ \exists S \subseteq \Sigma_n : \epsilon \leq \pi_g(S) \leq 1 - \epsilon,\ \mathcal{Q}_g(S, S^c) \leq \phi\}.\end{aligned} \tag{18.12}$$

Fix an ordering of the subsets of Σ_n, and let S_g be the first cut satisfying these conditions. The cut S_g depends only on g, and any fixed order works for the argument. The mass and flow of each cut are continuous finite sums in g, so this selection is measurable, as in the conventions of Section 13.1. The selected sets defined below are empty off the witness event.

Let

$$\epsilon_n = n^{-\mathsf{d}}, \qquad \phi_n = e^{-n^{1-\mathsf{q}}}, \qquad R_n = \lceil n^{1-\mathsf{r}} \rceil, \qquad A_n^{\mathrm{fld}} = n^{1-\mathsf{u}_{\mathrm{F}}}.$$

We write $d_H(x, y) = \#\{i \in [n] : x_i \neq y_i\} = \|x - y\|_2^2/4$ for the Hamming distance between $x, y \in \Sigma_n$, and $d_H(x, A) = \min_{y \in A} d_H(x, y)$, with $d_H(x, \varnothing) = +\infty$. We write ∂S for the vertex boundary that contains both endpoints of every edge across the cut, and define $\partial_R S = \{x : d_H(x, \partial S) \leq R\}$. Let $\mathcal{V}_R$ be the family of differences of the indicator vectors of two subsets of $[n]$ with at most R elements each. Then $v_i \in \{-1, 0, 1\}$ and

$\|v\|_1 = \|v\|_2^2 \le 2R$. The family contains both signs of every indicator of at most R sites, and

$$|\mathcal{V}_R| \le [(R+1)(n+1)^R]^2. \tag{18.13}$$

Define

$$\mathrm{Bad}_{R,A}(g) = \Big\{x : \exists v \in \mathcal{V}_R,\ \sum_i v_i (W(g)x)_i \ge A\Big\}.$$

The excluded set is

$$\mathcal{E}_n(g) = \partial_{R_n} S_g \ \cup\ \mathrm{Bad}_{R_n, A_n^{\mathrm{fld}}}(g) \quad \text{on } \mathcal{W}_n(\epsilon_n, \phi_n), \tag{18.14}$$

and is empty off that event. It contains the enlarged boundary of the selected cut and the spins with large sums of local fields.

Lemma 18.4. *Suppose that*

$$0 < \mathsf{q} < \mathsf{u}_{\mathrm{F}} < \mathsf{r} < 1, \qquad 2\mathsf{u}_{\mathrm{F}} - \mathsf{r} < \mathsf{q}, \qquad \mathsf{q} < 1/2. \tag{18.15}$$

Then, eventually,

$$\mathbb{E}_{\gamma_n} \pi_g(\mathcal{E}_n(g)) \le e^{-\frac{1}{2} n^{1-\mathsf{q}}}. \tag{18.16}$$

Proof. We first bound the annealed mass of $\mathrm{Bad}_{R,A}$ by conditioning on the spin. We then compare Gibbs weights along a short path from an enlarged boundary vertex to the boundary.

Fix a spin x. Recall that, under the joint law Π_n, the conditional density of the disorder given this spin is (18.1). For fixed v, the linear functional $f_v(g) = \sum_i v_i (W(g)x)_i$ has coefficients $a_{ij} = (v_i x_j + v_j x_i)/\sqrt{2n}$, so

$$\|a\|_2^2 = \|v\|_2^2 + (\textstyle\sum_i v_i x_i)^2/n \le 2\|v\|_2^2 \le 4R.$$

By integration by parts under (18.1), we have $|\mathbb{E}_{p_x} f_v| \le 2\beta \|v\|_1 \le 4\beta R$. Indeed, the derivative of $\beta H_g^c(x) - F^c(g)$ in the (i,j) entry is $\beta(x_i x_j - \langle y_i y_j \rangle_g)/\sqrt{2n}$, whose absolute value is at most $2\beta/\sqrt{2n}$. The density has Gaussian tails up to a linear exponential factor; by truncation, this justifies the integration by parts, including its boundary terms. Using (18.3) and a union bound over $\mathcal{V}_R$, we find that, for $A \ge 4\beta R$,

$$\mathbb{E}_{\gamma_n} \pi_g(\mathrm{Bad}_{R,A}(g)) \le [(R+1)(n+1)^R]^2 \exp\{-(A - 4\beta R)^2/(8R)\}. \tag{18.17}$$

When $R = R_n$ and $A = A_n^{\mathrm{fld}}$, the inequality $\mathsf{u}_{\mathrm{F}} < \mathsf{r}$ implies that $A \ge 8\beta R$ eventually. We also have that the logarithm of (18.13) is $o(A^2/R)$. Then (18.17) implies the upper bound $\exp(-c_\beta n^{1+\mathsf{r}-2\mathsf{u}_{\mathrm{F}}})$ eventually. By (18.15), the power of n in this exponent is strictly larger than $1 - \mathsf{q}$.

Next, fix a witness matrix and its selected cut. On $\|W(g)\|_{\mathrm{op}} \le 12$, we find by polarization that the change in energy when one spin is flipped is at most $24\sqrt{n}$. Then

every flip rate is at least $(1/2)e^{-24\beta\sqrt{n}}$. By reversibility, the flow in each direction across the selected cut is at most ϕ_n. We conclude that its two-sided boundary has Gibbs mass at most $4\phi_n e^{24\beta\sqrt{n}}$. If $x \notin \mathrm{Bad}_{R,A}(g)$ and $d_H(x,y) \le R$, expansion about x gives

$$|H_g^c(y) - H_g^c(x)| \le 2A + 24R. \tag{18.18}$$

To see this, let I be the set of flipped sites. Both $-x\mathbf{1}_I$ and its negative belong to $\mathcal{V}_R$, since $|I| \le R$. Their linear fields are bounded by A in both signs, and the quadratic correction is at most $2\|W(g)\|_{\mathrm{op}}R$. For each vertex of the enlarged boundary, we select a nearest boundary vertex, using a fixed order to break ties. By the bound on the size of a Hamming ball, each boundary vertex is selected at most $(R+1)(n+1)^R$ times. Comparing Gibbs weights by (18.18), we obtain

$$\begin{aligned}\pi_g(\mathcal{E}_n(g)) \le{}& 4(R_n+1)(n+1)^{R_n}\phi_n e^{\beta(2A_n^{\mathrm{fld}}+24R_n+24\sqrt{n})} \\ &+ \pi_g(\mathrm{Bad}_{R_n,A_n^{\mathrm{fld}}}(g)).\end{aligned} \tag{18.19}$$

The inequality also holds off the witness event, since its left side vanishes there. In the first term, the logarithms of the positive factors other than ϕ_n have powers $1-\mathsf{r}$ (with a factor $\log n$), $1-\mathsf{u}_{\mathrm{F}}$, and $1/2$, all smaller than $1-\mathsf{q}$. Averaging (18.19) and using (18.17), we obtain (18.16) after absorbing the sum and constants. □

We impose a pressure cutoff only at the outer target. We define

$$\begin{aligned}\mathcal{H}_n &= \{g : F^c(g) \le nP_\beta + n^{1-\mathsf{z}_{\mathrm{P}}}\}, \\ \overline{\mathcal{E}}_n(g) &= \begin{cases}\mathcal{E}_n(g), & g \in \mathcal{H}_n, \\ \varnothing, & g \notin \mathcal{H}_n.\end{cases}\end{aligned} \tag{18.20}$$

The gradient of F^c has norm at most $\beta\sqrt{n/2}$. By Gaussian concentration and $\mathbb{E}F^c \le nP_\beta$ (Proposition 4.21, Theorem 4.4, and Lemma 3.1), we have

$$\gamma_n(\mathcal{H}_n^c) \le e^{-c_\beta n^{1-2\mathsf{z}_{\mathrm{P}}}} = o(1) \qquad (0 < \mathsf{z}_{\mathrm{P}} < 1/2). \tag{18.21}$$

This is also Lemma 13.3. Each set is measurable by the conventions of Section 13.1; the disorder used in each coarse proposal is not restricted by $\mathcal{H}_n$.

18.3 The connecting sequence

We construct the sequence on one product space, following [85, Section 6, Proposition 7]. At the two endpoints we continue the coarse proposal to the fine horizon. At every other vertex we use only the coarse proposal. The simultaneous Gaussian rotation, continuity

of the query recursion, and pinning of spins to their magnetizations are the connecting-sequence method of that proof; we retain the source and replay histories on the growing mesh.

For the common fine schedule, let μ_n^{aux} be the independent product law of all D_i Wiener paths to V_{f}, auxiliary B_i^0 paths, used bridge paths through N_{f}, and full completion vectors $C_j \sim N(0, I_n)$ through N_{f}. We may retain the unused bridge coordinates independently; this gives the same finite experiment. At both horizons, the matrix has dimension n and the initial frame is the identity. By Lemma 14.1, these are the replay inputs.

Under a single product probability $\mathbb{P}_n$, we sample independently

$$g_\star, g_\circ, g_1, g_2 \sim \gamma_n, \qquad a_1, a_2 \sim \mu_n^{\mathrm{aux}},$$

and independent arrays of uniform thresholds at each grid point, together with two arrays for the fine endpoints. For $0 \le \theta \le \pi/2$, we define

$$\begin{aligned} g_{\mathrm{base}} &= \sqrt{1-\delta}\, g_\star + \sqrt{\delta}\, g_\circ, \\ g_\theta &= \sqrt{1-\delta}\, g_{\mathrm{base}} + \sqrt{\delta}(\cos\theta\, g_1 + \sin\theta\, g_2), \\ a_\theta &= \cos\theta\, a_1 + \sin\theta\, a_2. \end{aligned} \tag{18.22}$$

The rotation of the auxiliary inputs is defined by blockwise addition of vectors and paths, which preserves the zero endpoints of bridges. The uniform thresholds are not rotated. We use the coarse replay at every angle and continue it to the fine horizon at the two endpoints. Let

$$\ell_n = \lceil (\pi/2) e^{n^{\mathrm{grot}}} \rceil, \qquad \theta_j = \frac{\pi j}{2\ell_n} \quad (0 \le j \le \ell_n).$$

We denote the coarse spins by X_j^B, their coarse magnetizations by M_j, and the fine endpoint spins by X_0^F, X_1^F. For a grid angle $\theta = \theta_j$, we write $X_\theta^B = X_j^B$, and let U_j^B be its coarse threshold array. We define

$$E_{B,\theta_j} = \{(g_{\theta_j}, a_{\theta_j}|_B, U_j^B) \in E_B\},$$

where E_B is the event of Proposition 17.8 and $a|_B$ restricts the D paths to V_{b} and retains only the auxiliary cell, bridges, and completion vectors used by the coarse replay.

The outer pair $(g_\star, g_\circ)$ has law $\gamma_n^{\otimes 2}$. The internal law μ_n^{int} is the fixed product law of $(g_1, g_2, a_1, a_2, \text{thresholds})$, so $\mathbb{P}_n = \gamma_n^{\otimes 2} \otimes \mu_n^{\mathrm{int}}$. After an outer pair is fixed, the internal law remains the same. For each fixed angle, g_θ is standard Gaussian, and $a_\theta \sim \mu_n^{\mathrm{aux}}$ independently of all matrix drivers. For the complete paths, we note that every finite collection of path evaluations is jointly Gaussian, that the covariances within and between blocks are unchanged by the rotation, and that cylinders at rational times generate the path sigma algebra. After fixing $(g_\star, g_\circ)$, the spin X_0^F depends only on (g_1, a_1, U_0^F), and X_1^F depends only on (g_2, a_2, U_1^F). We have

$$\mathrm{Law}_{\mu_n^{\mathrm{int}}}(X_0^F, X_1^F) = \nu_{g_{\mathrm{base}}}^{\mathrm{sm}} \otimes \nu_{g_{\mathrm{base}}}^{\mathrm{sm}}. \tag{18.23}$$

This is the law of the endpoints under the internal measure, before any Gram, pinning, or crossing event is imposed.

Lemma 18.5. *If* $\mathsf{q} < \min(\mathsf{k_b}, \mathsf{z_P}, \mathsf{D})$*, then at each fixed grid angle, eventually,*

$$\mathbb{P}_n(E_{B,\theta},\ X_\theta^B \in \overline{\mathcal{E}}_n(g_\star)) \le e^{-\frac{1}{4}n^{1-\mathsf{q}}}. \tag{18.24}$$

Proof. We refine the target-to-proposal comparison in [85, Section 6, proof of Proposition 7] by integrating the conditional Gaussian energy at each fixed spin. The calculation below gives an $O_\beta(n\delta)$ regression cost while retaining the target Gibbs normalization. We condition only on $g_\star$. From (18.22), we have

$$\begin{aligned} g_\theta &= (1-\delta)g_\star + \sqrt{\delta(2-\delta)}\, Z_\theta, \\ Z_\theta &= \frac{\sqrt{1-\delta}\, g_\circ + \cos\theta\, g_1 + \sin\theta\, g_2}{\sqrt{2-\delta}}, \end{aligned} \tag{18.25}$$

where $Z_\theta \sim \gamma_n$ is independent of $g_\star$. For a fixed spin, $H_Z^c(x)$ is Gaussian with variance $n/2$, including the completed diagonal. It follows that

$$\mathbb{E}[e^{\beta H_{g_\theta}^c(x)} \mid g_\star] = e^{\beta(1-\delta)H_{g_\star}^c(x) + \beta^2 n\delta(2-\delta)/4}. \tag{18.26}$$

At a fixed angle, the auxiliary input and thresholds are independent of the entire matrix tuple. The restricted coarse kernel of Proposition 17.8 therefore still applies after adjoining $g_\star, g_\circ, g_\theta$. The conditional Gaussian in (18.25) has positive covariance, so it does not charge the common γ_n-null exceptional set of that proposition.

Recall that on the support of $\overline{\mathcal{E}}_n(g_\star)$, we have $\|W(g_\star)\|_{\mathrm{op}} \le 12$ and $|H_{g_\star}^c(x)| \le 6n$. Using Proposition 17.8 and (18.26), then summing over the set $\overline{\mathcal{E}}_n(g_\star)$ fixed by the target, we bound the conditional probability of hitting this set by

$$\exp\{n^{1-\mathsf{k_b}} + F^c(g_\star) - nP_\beta + (6\beta + \beta^2/2)n\delta\}\, \pi_{g_\star}(\overline{\mathcal{E}}_n(g_\star)).$$

Here we retained the target Gibbs normalization. By (18.20) and (18.16), its average in $g_\star$ is at most

$$\exp\{n^{1-\mathsf{k_b}} + n^{1-\mathsf{z_P}} + C_\beta n^{1-\mathsf{D}}\} e^{-\frac{1}{2}n^{1-\mathsf{q}}}.$$

By the strict inequalities between the exponents, this proves (18.24). The cost of the regression is of order $n\delta$; the exceptional event (18.21) for the target pressure is included only after the crossing argument in Section 18.4. □

Lemma 18.6. *There exists a measurable cap event, with failure probability at most* $C(N_{\mathsf{f}}+2)e^{-cn}$*, that controls all angles. On this event and all coarse Gram events, neighboring grid points satisfy*

$$\|M_{j+1} - M_j\| \le e^{C_\beta n^{\mathsf{a}} \log n} |\theta_{j+1} - \theta_j| \le e^{-n^{\mathsf{g_{rot}}}/2} \tag{18.27}$$

eventually, provided $\mathsf{a} < \mathsf{g_{rot}}$.

Proof. This is the input-cap and recursive-continuity argument of [85, Section 6, proof of Proposition 7], with the number of steps and the inverse scalar profiles controlled by the growing schedule. We first bound the primitive inputs uniformly in the angle. We then compare the two coarse recursions at neighboring grid points.

For a Wiener path of duration T, the reflection bound gives $\mathbb{P}(\sup|B| > u\sqrt{T}) \le 4e^{-u^2/2}$. Then, for some fixed small parameter, the exponential moment of $\sup|B|^2/T$ is bounded by a universal constant. By Chernoff's inequality over the independent rows, we have $\sum_i \sup|B_i|^2 \le CTn$ except with probability e^{-cn} when C is sufficiently large. A bridge has supremum at most twice that of its generating Wiener path. These bounds cover the two D blocks, the auxiliary cells, and the finitely many used bridge blocks. We also bound every coordinate of the four primitive matrices and of the two streams of completion vectors by n; this adds failure probability at most $C(n^2 + nN_{\mathsf{f}})e^{-n^2/2}$. All required path norms, completion norms, and matrix operator norms are then at most n^{C_0} for a fixed C_0; for example, an enlarged n^4 cap suffices. Rotated inputs satisfy the same cap up to a factor of two, and their differences are at most a polynomial times $|\theta - \phi|$. This bounds the inputs for all angles.

For the recursion, we use the path norm $(n^{-1}\sum_i \sup_t |b_i(t)|^2)^{1/2}$. On the target Gram events at the two grid points $\|A_{\mathcal{T}} - I\| \le 3/4$, every Gram–Schmidt residual has normalized norm at least $1/2$. Indeed, the squared norm of each residual is a Schur complement and is at least the minimum eigenvalue of the relevant Gram matrix. If u_l, u_l' are the earlier unit queries, the operator norm of their projection difference is at most $2\sum_l \|u_l - u_l'\|$. Normalization is 4-Lipschitz on vectors of norm at least $1/2$. The next query difference is bounded by a fixed multiple of the target difference and this sum of earlier differences.

Each target difference is bounded by $2s_j^{-1}$ times the scalar field difference on the preceding path. Using Picard's construction, Gronwall's inequality, and integration by parts for bounded-variation coefficients, we bound that difference by $Ce^{V_{\mathsf{f}}}$ times the gap between the inputs to the preceding drivers. All s_j^{-1} are polynomially bounded, since $s_j^2 \ge h/R_{V_{\mathsf{f}}}$ and Lemma 13.6 applies. By the query endpoint formula in Lemma 14.1, the next endpoint gap is bounded by a polynomial times the primitive matrix and completion gap plus the sum of preceding query gaps. Appending the next cell adds at most the old path gap, $\sqrt{h_j}$ times this endpoint gap, and the bridge gap. Here $\sqrt{h_j} \le 1$. The recursion implies that one chronological step multiplies the accumulated gap by at most n^{C_β} and adds at most that factor times the primitive input gap. After $N_{\mathsf{b}} \le 8n^{\mathsf{a}}$ steps, the amplification is at most $e^{C_\beta n^{\mathsf{a}} \log n}$. The final map m is spatially 1-Lipschitz. The grid spacing is at most $e^{-n^{\mathsf{g}_{\mathrm{rot}}}}$, proving (18.27). This comparison uses only coarse queries and the Gram events at the two endpoints of each grid edge. □

Lemma 18.7. *If* $\mathsf{r} < \mathsf{b}/2$, *each of the* $\ell_n + 3$ *pairs*

$$(X, M_B) = (X_j^B, M_j) \quad (0 \le j \le \ell_n),$$
$$(X_0^F, M_0), \qquad (X_1^F, M_{\ell_n})$$

satisfies

$$\mathbb{P}_n(\|X - M_B\|_2^2 > R_n/4) \le 2e^{-R_n/64} \tag{18.28}$$

eventually.

Proof. We use posterior projection and the independent-row pinning argument of [85, Section 6, proof of Proposition 7]. Under the unconditioned law at a single angle, the source identification in Section 14 gives the law of the complete coordinate rows, which are independent under this law. For $T = V_{\mathsf{b}}$ or V_{f}, posterior projection at $S = V_{\mathsf{b}}$ gives

$$\mathbb{E}[(X_T - m_S)^2] = 1 - \mathbb{E}m_S^2 = p(S) \le C_\beta n^{-\mathsf{b}/2}.$$

The squared distances lie in $[0, 4]$ and are independent across rows. Using $e^{z/4} \le 1 + (e-1)z/4$ for $0 \le z \le 4$ and Chernoff's inequality with parameter $1/4$, we find that, once $np(S) \le R_n/16$, their sum exceeds $R_n/4$ with probability at most $e^{-R_n/32}$. This implies (18.28). For the fine endpoints, the equality of the common initial segment after Lemma 14.3 gives the same m_S. We have proved the bounds under the unconditioned law of each run. □

Define the sequence

$$f_0 = X_0^F, \qquad f_{j+1} = X_j^B \quad (0 \le j \le \ell_n), \qquad f_{\ell_n+2} = X_1^F. \tag{18.29}$$

Every step has a coarse endpoint. Let Good_n be the intersection of the cap event, all E_{B,θ_j}, all pinning events in (18.28), and the avoidance events $X_j^B \notin \overline{\mathcal{E}}_n(g_\star)$ for every coarse vertex. This is a finite intersection of measurable events on the product space.

Proposition 18.8. *Assume the hypotheses of Proposition 17.8 and Lemmas 18.4 to 18.7. On* Good_n, *every step of* (18.29) *has Hamming length at most* R_n. *Moreover,*

$$\begin{aligned}\mathbb{P}_n(\mathrm{Good}_n^c) \le{}& C(N_{\mathsf{f}} + 2)e^{-cn} + (\ell_n + 1)e^{-n^{1-\mathsf{s}_{\mathsf{b}}}} \\ &+ 2(\ell_n + 3)e^{-R_n/64} + (\ell_n + 1)e^{-\frac{1}{4}n^{1-\mathsf{q}}}.\end{aligned} \tag{18.30}$$

If, in addition,

$$\begin{gathered}\mathsf{r} < \mathsf{s}_{\mathsf{b}}, \qquad \mathsf{q} < \mathsf{s}_{\mathsf{b}}, \qquad \mathsf{a} < \mathsf{g}_{\mathrm{rot}} < 1 - \mathsf{s}_{\mathsf{b}}, \\ \mathsf{s}_{\mathsf{b}} < \mathsf{s}_{\mathrm{seq}} < \mathsf{e},\end{gathered} \tag{18.31}$$

then, eventually,

$$\mathbb{P}_n(\mathrm{Good}_n^c) \le e^{-n^{1-\mathsf{s}_{\mathrm{seq}}}}. \tag{18.32}$$

Proof. For spins x, y and real vectors m, m', we have

$$4d_H(x,y) \le 3\{\|x-m\|_2^2 + \|m-m'\|_2^2 + \|m'-y\|_2^2\}.$$

Each term is at most $R_n/4$ on Good_n, eventually, by (18.27) and (18.28). For the steps between fine and coarse spins, we have that the middle term is zero, since the common initial segments agree.

A union bound using Proposition 17.8, (18.24), the cap bound, and (18.28) proves (18.30). The strict gaps in (18.31) absorb its constants and $\log(\ell_n+1) = O(n^{\mathsf{g_{rot}}})$, giving (18.32). The fine endpoints contribute their pinning failures to this union. The Gram and likelihood events concern only coarse proposals. Each bound holds under the unconditioned law of the corresponding run. This union bound does not use independence between different angles. □

18.4 Exclusion of low-flow witnesses

We now combine coverage with the connecting sequence. The crossing argument is the one in [85, Section 6, Proposition 7].

Proposition 18.9. *Assume the hypotheses of Lemmas 18.2 and 18.3 and Proposition 18.8, and let* $0 < \mathsf{z_P} < 1/2$. *Then*

$$\begin{aligned}\gamma_n(\mathcal{W}_n(\epsilon_n,\phi_n)) \le C_\beta\big(n^{\mathsf{u_E}-\mathsf{D}} + n^{\mathsf{u_E}-\mathsf{k_f}+\mathsf{D}}\big) + e^{-c_\beta n^{1-2\mathsf{z_P}}} \\ + e^{2n^{1-\mathsf{e}} - n^{1-\mathsf{s_{seq}}}} = o(1).\end{aligned} \tag{18.33}$$

Proof. Fix an outer pair $(g_\star, g_\circ) \in \mathcal{C}_n$ with $g_\star \in \mathcal{H}_n \cap \mathcal{W}_n(\epsilon_n,\phi_n)$. The selected cut $S_{g_\star}$ is ϵ_n-balanced. By (18.23) and Lemma 18.3, the event $f_0 \in S_{g_\star}$, $f_{\ell_n+2} \notin S_{g_\star}$ has internal probability at least ϖ_n.

If a sequence joins opposite sides with steps of Hamming length at most R_n, then some step has endpoints on opposite sides. A shortest path in the hypercube between the endpoints of that step contains an edge across the cut, and both endpoints of this edge lie in the boundary. Both endpoints of the step are within R_n of that boundary. One endpoint is coarse by (18.29), so it belongs to $\partial_{R_n} S_{g_\star}$, which is contained in $\mathcal{E}_n(g_\star) = \overline{\mathcal{E}}_n(g_\star)$ on this target. Crossing is therefore impossible on Good_n, and

$$\varpi_n \le \mu_n^{\mathrm{int}}\big(\mathrm{Good}_n(g_\star,g_\circ)^c\big). \tag{18.34}$$

The crossing event may depend on the fixed outer pair, while (18.23) gives the law of the endpoints under the unconditioned internal measure.

Integrating (18.34) against $\gamma_n^{\otimes 2}$ and using Fubini's theorem for the product measure, we obtain

$$\begin{aligned}\gamma_n(\mathcal{W}_n(\epsilon_n,\phi_n)) \le C_\beta\big(n^{\mathsf{u_E}-\mathsf{D}} + n^{\mathsf{u_E}-\mathsf{k_f}+\mathsf{D}}\big) + e^{-c_\beta n^{1-2\mathsf{z_P}}} \\ + \varpi_n^{-1}\mathbb{P}_n(\mathrm{Good}_n^c).\end{aligned}$$

Here we used Lemma 18.2 and (18.21) for the outer exceptions. By (18.11) and (18.32), this proves (18.33). The exceptional events for the pressure and for coverage each enter once, before we divide by the crossing probability. □

SECTION 19

Proof of the upper bound and the parameter system

We prove Theorem 1.3 and Corollaries 1.4 and 1.5 from Proposition 18.9 and Lemma 13.2. The finite-state argument of [85, Section 7, Lemma 8] converts the exclusion of balanced cuts with small flow into a bound on the average distance from equilibrium. We give its proof, then establish the estimate for the discrete chain. In Section 19.5, we check the parameter comparisons used throughout Part IV.

Throughout, the disorder, Gibbs measure, and continuous-time semigroup are those of Section 13.1. In particular, $\pi_g = \pi^{J(g)}$ and $P_t^g = P_t^{J(g)}$. The exponents are fixed before n tends to infinity.

19.1 From balanced-cut flow to mixing

The next lemma is [85, Lemma 8], specialized to the heat-bath chain. Its proof uses the coarea argument for the reversible Cheeger inequality of Lawler and Sokal [67, Theorem 3.5] and entropy dissipation as in Diaconis and Saloff-Coste [51, Lemmas 2.5 and 2.7]. We keep track of the estimates on the Dirichlet form, since they are also used in Section 19.3.

Lemma 19.1 (Pruning and averaged mixing)**.** *Fix $n \geq 2$, $\beta > 0$, and g. Let $0 < \epsilon < 1/3$ and $\phi > 0$. Suppose that*

$$\mathcal{Q}_g(S, S^c) > \phi \qquad \text{whenever } \epsilon \leq \pi_g(S) \leq 1 - \epsilon.$$

Then, for every $t > 0$,

$$\sum_{x \in \Sigma_n} \pi_g(x) \| P_t^g(x, \cdot) - \pi_g \|_{\mathrm{TV}} \leq \frac{n\sqrt{2 \log 2}}{\phi \sqrt{t}} + 2\epsilon. \tag{19.1}$$

Proof. We first remove sets that fail an internal expansion inequality. On the remaining set, we prove a Poincaré inequality by integrating over level sets. Entropy dissipation for the full chain then gives a time at which the average distance is small; contraction gives the estimate at t.

In this proof, we write $\pi = \pi_g$. For distinct configurations x, y, let $q(x,y) = \pi(x) p_i^{J(g)}(-x_i \mid x)$ if $y = x^{i,-x_i}$, and let $q(x,y) = 0$ otherwise. By reversibility, $q(x,y) = q(y,x)$, and

$$\sum_{y \neq x} q(x,y) \leq n\pi(x).$$

We use the Dirichlet form

$$\mathfrak{D}(u) = \sum_{\{x,y\}} q(x,y)(u(x) - u(y))^2 = -\langle u, \mathcal{L}^{J(g)} u \rangle_\pi, \tag{19.2}$$

where each unordered pair of distinct configurations occurs once and $\langle u, v \rangle_\pi = \sum_x \pi(x) u(x) v(x)$.

For disjoint sets $A, B \subseteq \Sigma_n$, we write

$$\mathcal{Q}_g(A,B) = \sum_{x \in A} \sum_{y \in B} q(x,y).$$

This agrees with (13.3) when $B = A^c$.

Removing sets. We use the pruning procedure of [85, Section 7, proof of Lemma 8]. We start with $C = \Sigma_n$. Whenever a nonempty set $U \subset C$ satisfies

$$\pi(U) \leq \pi(C)/2, \qquad \mathcal{Q}_g(U, C \setminus U) < \phi \pi(U),$$

we remove U from C. The procedure terminates after finitely many removals. Let A be the union of the sets removed up to any stage. Every edge from A to its current complement was counted when its first endpoint was removed. Summing the inequalities at those stages, we have, whenever A is nonempty,

$$\mathcal{Q}_g(A, \Sigma_n \setminus A) < \phi \pi(A).$$

Suppose that the removed mass reaches ϵ, and consider the first removal at which this happens. Before this removal, the removed mass was some $a < \epsilon$, and the newly removed set has mass at most $(1-a)/2$. After it, the removed mass is at most $(1+a)/2 < (1+\epsilon)/2 < 1 - \epsilon$. This contradicts the hypothesis on balanced cuts, because $\phi \pi(A) \leq \phi$. We have proved that the terminal set satisfies $\pi(C) > 1 - \epsilon$ and

$$\mathcal{Q}_g(U, C \setminus U) \geq \phi \pi(U) \quad \text{if } U \subset C \text{ and } \pi(U) \leq \pi(C)/2. \tag{19.3}$$

A Poincaré inequality on C. We apply the coarea proof of the reversible Cheeger bound [67, Theorem 3.5], with unnormalized weights on the retained set as in [85, Section 7, proof of Lemma 8]. We claim that, for every real function u on Σ_n,

$$\sum_{y \in C} \pi(y)(u(y) - \bar{u}_C)^2 \leq \frac{2n}{\phi^2} \mathfrak{D}(u), \qquad \bar{u}_C = \frac{1}{\pi(C)} \sum_{y \in C} \pi(y) u(y). \tag{19.4}$$

Let m be a weighted median of u on C. Each of $w = (u-m)_+$ and $w = (m-u)_+$ has support of mass at most $\pi(C)/2$. Integrating (19.3) over the level sets of w^2, we find

$$\phi \sum_C \pi w^2 \le \sum_{\{x,y\}\subset C} q(x,y)|w(x)^2 - w(y)^2|.$$

By Cauchy–Schwarz, the right side is at most

$$\left(\sum_{\{x,y\}\subset C} q(x,y)(w(x)-w(y))^2 \right)^{1/2} \left(2n \sum_C \pi w^2 \right)^{1/2}.$$

Here we used $(w(x)+w(y))^2 \le 2(w(x)^2 + w(y)^2)$ and the bound on the outgoing rates. Dividing and squaring, we find the desired estimate for w when $\sum_C \pi w^2 > 0$; when this sum vanishes, the estimate is immediate. The sum of the energies of the two parts is at most $\mathfrak{D}(u)$. Since the weighted mean minimizes the squared error, (19.4) follows.

Entropy dissipation. The derivative formula and square-root Dirichlet bound are [51, Lemmas 2.5 and 2.7], with the rates of Section 13.1. For a fixed starting configuration x, define $f_s^x(y) = P_s^g(x,y)/\pi(y)$. We write

$$\mathrm{Ent}_\pi(f) = \sum_y \pi(y) f(y) \log f(y), \qquad H(\pi) = -\sum_x \pi(x) \log \pi(x),$$

where $\pi(f) = 1$ and $0 \log 0 = 0$. Reversibility implies

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} \mathrm{Ent}_\pi(f_s^x) &= -\sum_{\{y,z\}} q(y,z)(f_s^x(y) - f_s^x(z))(\log f_s^x(y) - \log f_s^x(z)) \\ &\le -4\mathfrak{D}(\sqrt{f_s^x}). \end{aligned}$$

In the last step, we used $(a-b)(\log a - \log b) \ge 4(\sqrt{a} - \sqrt{b})^2$ for $a, b > 0$. The initial entropy is $\log(1/\pi(x))$, and the final entropy is nonnegative. Integrating and averaging over $x \sim \pi$, we find

$$\int_0^t \sum_x \pi(x) \mathfrak{D}(\sqrt{f_s^x})\, \mathrm{d}s \le \frac{1}{4} H(\pi). \tag{19.5}$$

To justify the calculation when the density has zero entries, we start with $(1-\eta) f_0^x + \eta$ and let $\eta \downarrow 0$. Since the state space is finite and entropy is continuous at zero, the integrated inequality passes to this limit.

Averaging the distances. By (19.4) and (19.5), there exists $s \in (0, t]$ such that

$$\begin{aligned} &\sum_x \pi(x) r_x^2 \le \frac{nH(\pi)}{2t\phi^2}, \\ &\qquad r_x^2 = \sum_{y\in C} \pi(y)(\sqrt{f_s^x(y)} - a_x)^2, \qquad a_x = \frac{1}{\pi(C)} \sum_{y\in C} \pi(y) \sqrt{f_s^x(y)}. \end{aligned} \tag{19.6}$$

Using Cauchy–Schwarz and the fact that $a_x^2 \pi(C) \le \sum_C \pi f_s^x \le 1$, we have

$$\sum_C \pi |f_s^x - a_x^2| \le 2 r_x.$$

Further, the normalization of the density gives $|\sum_C \pi(f_s^x - 1)| \le \sum_{C^c} \pi(f_s^x + 1)$. Comparing f_s^x first with a_x^2 and then with 1 on C, and adding the complement, we have

$$\sum_{\Sigma_n} \pi |f_s^x - 1| \le 4 r_x + 2 \sum_{C^c} \pi(f_s^x + 1). \tag{19.7}$$

By stationarity, we have $\sum_x \pi(x) f_s^x(y) = 1$. It follows that the last sum in the preceding display, without its factor two, has average $2\pi(C^c)$ over $x \sim \pi$. Dividing (19.7) by two and using Cauchy–Schwarz in x, we obtain

$$\sum_x \pi(x) \| P_s^g(x, \cdot) - \pi \|_{\mathrm{TV}} \le \frac{\sqrt{2nH(\pi)}}{\phi \sqrt{t}} + 2\epsilon.$$

Since $|\Sigma_n| = 2^n$, we have $H(\pi) \le n \log 2$. Total variation contracts under P_{t-s}^g by Lemma 3.2, which gives (19.1) at time t. □

19.2 Proof of the main theorem

We now combine the witness estimate with Lemma 19.1. Recall that $\epsilon_n = n^{-\mathsf{d}}$, $\phi_n = e^{-n^{1-\mathsf{q}}}$, and $\mathcal{W}_n(\epsilon_n, \phi_n)$ is the event that $\|W(g)\|_{\mathrm{op}} \le 12$ and some ϵ_n-balanced cut has flow at most ϕ_n.

Proof of Theorem 1.3. Fix $\beta > 1$ and $0 < \omega < 1/4$, and take the affine family of Section 13.3 with $\varepsilon_\omega = (1/4 - \omega)/20$. The comparisons verified in Section 19.5 imply all hypotheses of the preceding estimates. By Proposition 18.9, we have

$$\gamma_n(\mathcal{W}_n(\epsilon_n, \phi_n)) \le C_\beta n^{-\varepsilon_\omega} + e^{-c_\beta n^{2/5}} + \exp\{2n^{3/4+3\varepsilon_\omega} - n^{3/4+4\varepsilon_\omega}\} = o(1). \tag{19.8}$$

Outside this event and $\{\|W(g)\|_{\mathrm{op}} > 12\}$, every ϵ_n-balanced cut has flow greater than ϕ_n. For all sufficiently large n, we have $0 < \epsilon_n < 1/3$. We use (19.1) with

$$\epsilon = \epsilon_n, \qquad \phi = \phi_n, \qquad t = e^{n^{1-\omega}}.$$

Using Markov's inequality in the Gibbs draw and Lemma 13.2, we find

$$\begin{aligned} &\Pi_n\big\{(g, x) : \| P^g_{e^{n^{1-\omega}}}(x, \cdot) - \pi_g \|_{\mathrm{TV}} > 1/4\big\} \\ &\quad \le \gamma_n(\mathcal{W}_n(\epsilon_n, \phi_n)) + 2e^{-n} + 4n\sqrt{2\log 2}\, e^{n^{1-\mathsf{q}} - \frac{1}{2} n^{1-\omega}} + 8n^{-\mathsf{d}}. \end{aligned} \tag{19.9}$$

The right side tends to zero, since $\mathsf{d} > 0$ and $\mathsf{q} - \omega = 10\varepsilon_\omega > 0$. By the dictionary in Section 13.1, the $(J(g), x)$-marginal of Π_n is $\mathbb{P}_{J, X_0 \sim \pi^J}$. This proves the assertion under the joint law. Here the Gibbs draw is held fixed when the distance is formed, and the semigroup has rate one per site. □

19.3 Proof of the discrete-time corollary

We use the retained set and the Poincaré inequality from the proof of Lemma 19.1. What changes is the entropy dissipation inequality, which we prove for one step of the discrete heat-bath chain. The subsequent averaging argument is that of [85, Section 7, Lemma 8].

Proof of Corollary 1.4. Fix $\beta > 1$ and $0 < \omega < 1/4$, and choose the same exponents as in the proof of Theorem 1.3. We first fix a disorder g outside the exceptional events of that proof. We abbreviate $\pi = \pi_g$ and $K = K^{J(g)}$. For a function u on Σ_n, define

$$Q_i u(x) = \sum_{\sigma \in \{-1,1\}} p_i^{J(g)}(\sigma \mid x) u(x^{i,\sigma}).$$

Each Q_i is conditional expectation given all spins except the ith, and $K = n^{-1} \sum_i Q_i$. In $L^2(\pi)$, the operators Q_i are orthogonal projections. By (19.2),

$$\mathfrak{D}(u) = \sum_{i=1}^{n} \langle u, (I - Q_i) u \rangle_\pi. \tag{19.10}$$

Let $f \geq 0$ satisfy $\pi(f) = 1$. The convexity of $s \log s$ implies that $\mathrm{Ent}_\pi(Kf) \leq n^{-1} \sum_i \mathrm{Ent}_\pi(Q_i f)$. We claim that

$$\mathrm{Ent}_\pi(f) - \mathrm{Ent}_\pi(Q_i f) \geq \langle \sqrt{f}, (I - Q_i) \sqrt{f} \rangle_\pi.$$

Indeed, on each fiber obtained by fixing all spins except the ith, write $a = Q_i f$. For $s \geq 0$ and $a > 0$,

$$s \log(s/a) - s + a \geq (\sqrt{s} - \sqrt{a})^2.$$

This follows by writing the difference as $2a\sqrt{s/a}(\sqrt{s/a} \log \sqrt{s/a} - \sqrt{s/a} + 1) \geq 0$, with its continuous interpretation at $s = 0$. Averaging on the fiber eliminates $-s + a$. The resulting squared error is at least the squared error about the fiber mean $Q_i \sqrt{f}$. Fibers with $a = 0$ have $f = 0$ throughout and contribute zero. Averaging over the other spins proves the claim. Combining it with (19.10), we have

$$\mathrm{Ent}_\pi(f) - \mathrm{Ent}_\pi(Kf) \geq \frac{1}{n} \mathfrak{D}(\sqrt{f}). \tag{19.11}$$

For $f_k^x(y) = K^k(x, y)/\pi(y)$, reversibility gives $f_{k+1}^x = K f_k^x$. Telescoping (19.11) and averaging over $x \sim \pi$, we obtain, for every integer $m \geq 1$,

$$\sum_{k=0}^{m-1} \sum_x \pi(x) \mathfrak{D}(\sqrt{f_k^x}) \leq n H(\pi) \leq n^2 \log 2. \tag{19.12}$$

Since the disorder lies outside the exceptional events, the hypothesis on balanced cuts holds with $\phi = \phi_n$ and $\epsilon = \epsilon_n$, and it gives the same set C and inequality (19.4). Then there exists $k \in \{0, \dots, m-1\}$ such that

$$\sum_x \pi(x) \sum_{y \in C} \pi(y) \bigl(\sqrt{f_k^x(y)} - a_x\bigr)^2 \le \frac{2n^2 H(\pi)}{m \phi_n^2}, \qquad a_x = \frac{1}{\pi(C)} \sum_{y \in C} \pi(y) \sqrt{f_k^x(y)}.$$

Using (19.7), stationarity, and Cauchy–Schwarz, we find

$$\sum_x \pi(x) d_{\mathrm{disc}}^{J(g)}(x, k) \le \frac{2n\sqrt{2H(\pi)}}{\phi_n \sqrt{m}} + 2\epsilon_n.$$

Using that the distance contracts under K^{m-k}, we find

$$\sum_x \pi_g(x) d_{\mathrm{disc}}^{J(g)}(x, m) \le \frac{2\sqrt{2 \log 2}\, n^{3/2} e^{n^{1-\mathsf{q}}}}{\sqrt{m}} + 2n^{-\mathsf{d}}. \tag{19.13}$$

Let $m = \lfloor e^{n^{1-\omega}} \rfloor$. For all sufficiently large n, we have $m \ge e^{n^{1-\omega}}/2$. The first term in (19.13) is at most

$$C n^{3/2} \exp\{n^{1-\mathsf{q}} - \tfrac{1}{2} n^{1-\omega}\},$$

which tends to zero because $\omega < \mathsf{q}$. By Markov's inequality in the Gibbs draw, we obtain

$$\begin{aligned} &\Pi_n\{(g, x) : d_{\mathrm{disc}}^{J(g)}(x, m) > 1/4\} \\ &\quad \le \gamma_n(\mathcal{W}_n(\epsilon_n, \phi_n)) + 2e^{-n} + C n^{3/2} e^{n^{1-\mathsf{q}} - \frac{1}{2} n^{1-\omega}} + 8n^{-\mathsf{d}} = o(1). \end{aligned} \tag{19.14}$$

Using the marginal identity for Π_n, we conclude that Corollary 1.4 holds. □

19.4 Proof of the two-sided corollary

The two-sided statement uses the lower bound at threshold $1/4$ and the two upper bounds just proved.

Proof of Corollary 1.5. Fix $\beta > 1$ and $0 < \omega < 1/4$, and let A_β be the coefficient in Theorem 1.1. With $\tau = 1/4$, that theorem implies

$$\mathbb{P}_{J, X_0 \sim \pi^J}\{d^J(X_0, e^{A_\beta n^{7/30}}) > 1/4\} \longrightarrow 1.$$

By the definition of the mixing time in (3.14) and Lemma 3.2, this event is $\{T^J(X_0) > e^{A_\beta n^{7/30}}\}$. Intersecting it with the event of probability tending to one in Theorem 1.3 proves the continuous-time assertion.

For the discrete chain, by (1.3) at $\tau = 1/4$, we have, with probability tending to one,

$$d^J_{\mathrm{disc}}(X_0, \lfloor e^{A_\beta n^{7/30}} \rfloor) > 1/4.$$

Since $k \mapsto d^J_{\mathrm{disc}}(X_0, k)$ is nonincreasing, this implies $T^J_{\mathrm{disc}}(X_0) > \lfloor e^{A_\beta n^{7/30}} \rfloor$. On the event in Corollary 1.4, we also have $T^J_{\mathrm{disc}}(X_0) \le \lfloor e^{n^{1-\omega}} \rfloor$. The probability of their intersection tends to one, as desired. □

19.5 The parameter system

We verify all comparisons used in the preceding chapters. Fix $0 < \omega < 1/4$ and define $\varepsilon_\omega = (1/4 - \omega)/20$. Then $0 < \varepsilon_\omega < 1/80$ and $\omega = 1/4 - 20\varepsilon_\omega$. The affine family from Section 13.3 is

$$\begin{aligned}
\mathsf{a} &= \tfrac{1}{2}, & \mathsf{f} &= \tfrac{1}{2}, & \mathsf{k_f} &= \tfrac{1}{2} - 2\varepsilon_\omega,\\
\mathsf{D} &= \tfrac{1}{4} - \varepsilon_\omega, & \mathsf{u_E} &= \tfrac{1}{4} - 2\varepsilon_\omega, & \mathsf{d} &= \tfrac{1}{2}\varepsilon_\omega,\\
\mathsf{e} &= \tfrac{1}{4} - 3\varepsilon_\omega, & \mathsf{b} &= \tfrac{1}{2} - 14\varepsilon_\omega, & \mathsf{z} &= \tfrac{1}{4} - 8\varepsilon_\omega,\\
\mathsf{w} &= \tfrac{1}{8} - 3\varepsilon_\omega, & \mathsf{k_b} &= \tfrac{1}{4} - 9\varepsilon_\omega, & \mathsf{s_b} &= \tfrac{1}{4} - 5\varepsilon_\omega,\\
\mathsf{s_{seq}} &= \tfrac{1}{4} - 4\varepsilon_\omega, & \mathsf{r} &= \tfrac{1}{4} - \tfrac{15}{2}\varepsilon_\omega, & \mathsf{u_F} &= \tfrac{1}{4} - \tfrac{39}{4}\varepsilon_\omega,\\
\mathsf{q} &= \tfrac{1}{4} - 10\varepsilon_\omega, & \mathsf{z_P} &= \tfrac{3}{10}, & \mathsf{g_{rot}} &= \tfrac{3}{5}.
\end{aligned} \tag{19.15}$$

An optional intermediate loss $\mathsf{s_F}$ can be chosen to satisfy $2\mathsf{u_F} - \mathsf{r} < \mathsf{s_F} < \mathsf{q}$. We leave it unspecified; only the strict comparison $2\mathsf{u_F} - \mathsf{r} < \mathsf{q}$ enters the probability estimates.

The sufficient system consists of the positivity of the exponents and

$$\begin{gathered}
\mathsf{a} > \mathsf{f}/4, \qquad \mathsf{a} > \mathsf{b}/4, \qquad \mathsf{a} + \mathsf{f}/2 < 1,\\
\mathsf{k_f} < \min(\mathsf{a}, \mathsf{f}, 1 - \mathsf{a}), \qquad \mathsf{b} < \mathsf{f},\\
0 < \mathsf{k_b} < \mathsf{z} < \min(\mathsf{a}, \mathsf{b}/2, 2\mathsf{w}),\\
\max\{2\mathsf{w}, \mathsf{w} + \mathsf{b}/4, \mathsf{b}/2, 2\mathsf{z} - \mathsf{a}, 2\mathsf{z} - 2\mathsf{w}, 2\mathsf{z} - \mathsf{b}\} < \mathsf{s_b},\\
\mathsf{a} < 1 - \mathsf{s_b}, \qquad \mathsf{e} + \mathsf{d} < \mathsf{u_E} < \min(\mathsf{D}, \mathsf{k_f} - \mathsf{D}),\\
\mathsf{q} < \mathsf{u_F} < \mathsf{r} < \mathsf{b}/2, \qquad 2\mathsf{u_F} - \mathsf{r} < \mathsf{q}, \qquad \mathsf{q} < 1/2,\\
\mathsf{q} < \min(\mathsf{k_b}, \mathsf{D}, \mathsf{z_P}), \qquad 0 < \mathsf{z_P} < 1/2,\\
\mathsf{q} < \mathsf{s_b}, \qquad \mathsf{r} < \mathsf{s_b}, \qquad \mathsf{a} < \mathsf{g_{rot}} < 1 - \mathsf{s_b},\\
\mathsf{s_b} < \mathsf{s_{seq}} < \mathsf{e}, \qquad \omega < \mathsf{q}.
\end{gathered} \tag{19.16}$$

All exponents are below one whenever a growing cutoff or a tail estimate requires this. The scalar, Gram, concentration, regression, and finite-chain arguments show that these comparisons suffice.

We list the powers used in those arguments in Table 19.1. A rate is the power in a negative logarithm. Polylogarithmic denominators in rates are retained until a

strict comparison absorbs them. Query counts and expectation bounds may have fixed polylogarithmic numerators. The last column gives the instance $\omega = 1/5$, $\varepsilon_\omega = 1/400$.

Table 19.1. Powers in the parameter system.

Quantity	Family power	$\omega = 1/5$
Query count $N_{\mathsf{f}}, N_{\mathsf{b}}$	$1/2$	$1/2$
Fine maximum-cell saving	$3/8$	$3/8$
Coarse maximum-cell saving	$3/8 + 7\varepsilon_\omega/2$	$307/800$
Fine nh, $ne^{-V_{\mathsf{f}}}$, $N_{\mathsf{f}}(1+V_{\mathsf{f}})$	$1/2$	$1/2$
Fine $\sqrt{N_{\mathsf{f}}}$, $N_{\mathsf{f}}^2 V_{\mathsf{f}} R_{V_{\mathsf{f}}}/n$	$1/4$	$1/4$
Fine KL cutoff $1-\mathsf{k}_{\mathsf{f}}$	$1/2 + 2\varepsilon_\omega$	$101/200$
Weighted variance loss $2\mathsf{w}$	$1/4 - 6\varepsilon_\omega$	$47/200$
Weighted scale loss $\mathsf{w} + \mathsf{b}/4$	$1/4 - 13\varepsilon_\omega/2$	$187/800$
Fixed Gram loss $\mathsf{b}/2$	$1/4 - 7\varepsilon_\omega$	$93/400$
Weighted raw rate $1-2\mathsf{w}$	$3/4 + 6\varepsilon_\omega$	$153/200$
Weighted scale raw rate $1-\mathsf{w}-\mathsf{b}/4$	$3/4 + 13\varepsilon_\omega/2$	$613/800$
Fixed Gram raw rate $1-\mathsf{b}/2$	$3/4 + 7\varepsilon_\omega$	$307/400$
Joint lag / diagonal base rate $1+\mathsf{a}-2\mathsf{z}$	$1 + 16\varepsilon_\omega$	$26/25$
Gaussian coefficient rate $1+2\mathsf{w}-2\mathsf{z}$	$3/4 + 10\varepsilon_\omega$	$31/40$
Gaussian terminal rate $1+\mathsf{b}-2\mathsf{z}$	$1 + 2\varepsilon_\omega$	$201/200$
Coarse one-run rate $1-\mathsf{s}_{\mathsf{b}}$	$3/4 + 5\varepsilon_\omega$	$61/80$
Weighted quadratic energy saving $2\mathsf{w}$	$1/4 - 6\varepsilon_\omega$	$47/200$
Coarse log-density cost $1-\mathsf{k}_{\mathsf{b}}$	$3/4 + 9\varepsilon_\omega$	$309/400$
Both expected reverse-KL costs	$3/4 + \varepsilon_\omega$	$301/400$
Outer reverse-KL cutoff $1-\mathsf{u}_{\mathrm{E}}$	$3/4 + 2\varepsilon_\omega$	$151/200$
Both outer Markov decay savings	ε_ω	$1/400$
Raw crossing exponent $1-\mathsf{u}_{\mathrm{E}}+\mathsf{d}$	$3/4 + 5\varepsilon_\omega/2$	$121/160$
Crossing envelope $1-\mathsf{e}$	$3/4 + 3\varepsilon_\omega$	$303/400$
Total sequence failure rate $1-\mathsf{s}_{\mathrm{seq}}$	$3/4 + 4\varepsilon_\omega$	$19/25$
Pinning/radius power $1-\mathsf{r}$	$3/4 + 15\varepsilon_\omega/2$	$123/160$
Flow / good-output hit rate $1-\mathsf{q}$	$3/4 + 10\varepsilon_\omega$	$31/40$
Sparse-field threshold $1-\mathsf{u}_{\mathrm{F}}$	$3/4 + 39\varepsilon_\omega/4$	$1239/1600$
Hamming-ball entropy power $1-\mathsf{r}$	$3/4 + 15\varepsilon_\omega/2$	$123/160$
Bad-field rate $1+\mathsf{r}-2\mathsf{u}_{\mathrm{F}}$	$3/4 + 12\varepsilon_\omega$	$39/50$
Intermediate bad-field rate (optional)	see text	$(31/40, 39/50)$
Target pressure cutoff / rate	$7/10,\ 2/5$	$7/10,\ 2/5$
Regression log-cost $1-\mathsf{D}$	$3/4 + \varepsilon_\omega$	$301/400$
Rotation logarithmic grid size	$3/5$	$3/5$
Physical time exponent $1-\omega$	$3/4 + 20\varepsilon_\omega$	$4/5$

In the optional row, the last entry is the interval of allowed values. In the family, that interval is $(3/4 + 10\varepsilon_\omega, 3/4 + 12\varepsilon_\omega)$.

For example, $\max\{2\mathsf{w}, \mathsf{w} + \mathsf{b}/4\} = 2\mathsf{w}$, since the difference of the two entries is $\varepsilon_\omega/2$. The weighted variance rate and the fixed-Gram rate exceed the one-run rate by ε_ω and $2\varepsilon_\omega$, respectively. Each of these rates exceeds the net exponent $\mathsf{a} = 1/2$ with a fixed positive gap. We first establish the coarse event with rate $1 - \mathsf{s}_\mathsf{b}$, then take a union over the rotation grid. Since $\mathsf{s}_{\mathrm{seq}} - \mathsf{s}_\mathsf{b} = \varepsilon_\omega$ and $\mathsf{e} - \mathsf{s}_{\mathrm{seq}} = \varepsilon_\omega$, the grid union and the comparison with the rare crossing each retain a strict gap. The target pressure failure, whose rate is $2/5$, is charged once outside that comparison.

The parameter instance at $\omega = 1/5$ is

$$\begin{array}{llll} \omega = \frac{1}{5}, & \mathsf{q} = \frac{9}{40}, & \mathsf{a} = \frac{1}{2}, & \mathsf{f} = \frac{1}{2}, \\ \mathsf{k}_\mathsf{f} = \frac{99}{200}, & \mathsf{D} = \frac{99}{400}, & \mathsf{u}_\mathsf{E} = \frac{49}{200}, & \mathsf{d} = \frac{1}{800}, \\ \mathsf{e} = \frac{97}{400}, & \mathsf{b} = \frac{93}{200}, & \mathsf{z} = \frac{23}{100}, & \mathsf{w} = \frac{47}{400}, \\ \mathsf{k}_\mathsf{b} = \frac{91}{400}, & \mathsf{s}_\mathsf{b} = \frac{19}{80}, & \mathsf{s}_{\mathrm{seq}} = \frac{6}{25}, & \mathsf{r} = \frac{37}{160}, \\ \mathsf{u}_\mathsf{F} = \frac{361}{1600}, & \mathsf{z}_\mathsf{P} = \frac{3}{10}, & \mathsf{g}_{\mathrm{rot}} = \frac{3}{5}. & \end{array} \tag{19.17}$$

Some of the tight comparisons are listed below; the slack is the right side minus the left side.

Comparison	Family slack	$\omega = 1/5$
$\mathsf{k}_\mathsf{f} < \mathsf{a}$	$2\varepsilon_\omega$	$1/200$
$\mathsf{u}_\mathsf{E} < \mathsf{D}$, $\mathsf{u}_\mathsf{E} < \mathsf{k}_\mathsf{f} - \mathsf{D}$	ε_ω	$1/400$
$\mathsf{e} + \mathsf{d} < \mathsf{u}_\mathsf{E}$	$\varepsilon_\omega/2$	$1/800$
$\mathsf{z} < \mathsf{b}/2$	ε_ω	$1/400$
$\mathsf{z} < 2\mathsf{w}$	$2\varepsilon_\omega$	$1/200$
$2\mathsf{w} < \mathsf{s}_\mathsf{b}$	ε_ω	$1/400$
$\mathsf{w} + \mathsf{b}/4 < \mathsf{s}_\mathsf{b}$	$3\varepsilon_\omega/2$	$3/800$
$\mathsf{b}/2 < \mathsf{s}_\mathsf{b}$	$2\varepsilon_\omega$	$1/200$
$2\mathsf{z} - 2\mathsf{w} < \mathsf{s}_\mathsf{b}$	$5\varepsilon_\omega$	$1/80$
$\mathsf{q} < \mathsf{k}_\mathsf{b}$	ε_ω	$1/400$
$\mathsf{q} < \mathsf{u}_\mathsf{F}$	$\varepsilon_\omega/4$	$1/1600$
$\mathsf{u}_\mathsf{F} < \mathsf{r}$	$9\varepsilon_\omega/4$	$9/1600$
$\mathsf{r} < \mathsf{b}/2$	$\varepsilon_\omega/2$	$1/800$
$2\mathsf{u}_\mathsf{F} - \mathsf{r} < \mathsf{q}$	$2\varepsilon_\omega$	$1/200$
$\mathsf{s}_\mathsf{b} < \mathsf{s}_{\mathrm{seq}}$, $\mathsf{s}_{\mathrm{seq}} < \mathsf{e}$	ε_ω	$1/400$
$\omega < \mathsf{q}$	$10\varepsilon_\omega$	$1/40$

The least slack in the full comparison list at this instance is $1/1600$. The witness estimate specializes to

$$\gamma_n(\mathcal{W}_n(\epsilon_n, \phi_n)) \le C_\beta n^{-1/400} + e^{-c_\beta n^{2/5}} + \exp\{2n^{303/400} - n^{19/25}\} = o(1). \tag{19.19}$$

The bound (19.9) further adds the norm exception $2e^{-n}$, the term $4n\sqrt{2\log 2}\exp\{n^{31/40} - n^{4/5}/2\}$, and $8n^{-1/800}$.

Proposition 19.2 (Every fixed saving below one quarter)**.** *The family* (19.15) *satisfies every sufficient comparison in* (19.16) *for every* $0 < \varepsilon_\omega < 1/80$*. Hence Theorem 1.3 holds for every fixed* $0 < \omega < 1/4$*. These estimates do not reach* $\omega = 1/4$*.*

Proof. All parameters and all differences between the right and left sides of (19.16) are affine in ε_ω, after expanding each maximum and minimum into its constituent comparisons. Using (19.15), we find that all endpoint values at 0 and $1/80$ are nonnegative and that no difference is identically zero. Each interior value is a strict convex combination of these two endpoint values, at least one of which is positive. Every comparison is therefore strict. This also proves that $\omega = 1/4 - 20\varepsilon_\omega$ is positive; zero values at the excluded endpoints are allowed. The optional s_F can be chosen between $2u_F - r$ and q, whose difference is $2\varepsilon_\omega$.

In particular, $q - \omega = 10\varepsilon_\omega > 0$. By Proposition 18.9 and (19.9), we conclude that the asserted range holds. All constants are chosen after ω and β are fixed.

We now explain the obstruction in these estimates. Fine entropy requires $k_f < \min(a, 1-a) \le 1/2$. Reverse smoothing and coverage require

$$e + d < u_E < \min(D, k_f - D) \le k_f/2 < 1/4.$$

The connecting sequence needs $q < s_b < s_{seq} < e$, and mixing needs $\omega < q$. For any parameters satisfying these strict estimates, we have $\omega < 1/4$, independently of the affine family. At $\varepsilon_\omega = 0$, the fine, coverage, concentration, and crossing gaps vanish. Constants and logarithms cannot be absorbed by the arguments above. This is an obstruction to these estimates; it does not give a lower bound or an optimality statement for the SK dynamics. □

A calculation in exact rational arithmetic checks the comparisons; it lists each affine slack and its values at 0, $1/80$, and $1/400$. With an optional intermediate loss chosen inside the stated interval, the calculation checks 114 strict comparisons and 37 identities for the powers in Table 19.1. The proof of Proposition 19.2 does not rely on this calculation: the endpoint argument there gives the comparisons on the whole open interval.

PART V

Fluctuations and mean bias

SECTION 20

The fluctuation scale

This section concerns Theorem 1.6, the fluctuation theorem of OpenAI [81, Theorem 1.1]. Its proof in [81] has two parts of very different lengths. The main part, which occupies Sections 2–7.4 of that paper, is a two-sided bound on a positive logarithmic moment of the completed log partition function at a tilt slightly larger than $n^{-1/6}$. We state this bound in our normalization as Proposition 20.1, and in Section 20.2 we summarize how it is proved in [81]; we do not reprove it. The second part converts Proposition 20.1 into the three assertions of Theorem 1.6. It rests on an exponential augmentation argument of Aronow and Lopatto [8, Proposition 5.1], in the form given in [81, Lemmas 7.6 and 7.7 and Section 7.6]. We give complete proofs of this part: the log-concavity lemma (Lemma 20.2) and elementary bounds for one-dimensional log-concave laws (Lemma 20.3) in Section 20.3, and the deduction of Theorem 1.6 from Proposition 20.1 in Section 20.4. Section 21 uses the lower bound on the variance in Theorem 1.6.

As recalled in Section 1, OpenAI's companion preprint [82] claims more than Theorem 1.6: a positive limit of $n^{-1/3}\operatorname{Var}(F_n)$ and a nondegenerate limit law for $(F_n - \mathbb{E}F_n)/\operatorname{Var}(F_n)^{1/2}$ [82, Theorems 1.1 and 1.2]. We do not use these claims; their consequence for the mean bias is recalled in Remark 21.2.

Throughout this section, $\beta > 1$ is fixed, and constants may depend on β. Recall from Section 3 that the completed log partition function is $F_n^c = F_n + \beta G$, where G is a centered Gaussian variable of variance $1/2$, independent of the disorder J. For $t \in \mathbb{R}$, we define the logarithmic moment generating function of the centered variable $F_n^c - \mathbb{E}F_n^c$,

$$K_n(t) = \log \mathbb{E} \exp\bigl\{ t\bigl(F_n^c - \mathbb{E}F_n^c\bigr)\bigr\}. \tag{20.1}$$

By Lemma 3.1, $K_n(t)$ is finite for every $t \in \mathbb{R}$, F_n^c has mean $\mathbb{E}F_n$ and variance $\operatorname{Var}(F_n) + \beta^2/2$, and $K_n(t) = \log \mathbb{E}e^{t(F_n - \mathbb{E}F_n)} + \beta^2 t^2/4$ by (3.6). The proof of that lemma also shows that F_n has finite exponential moments, so all its moments are finite.

20.1 The two-sided bound on positive logarithmic moments

Proposition 20.1 (Two-sided bound on positive logarithmic moments)**.** *Fix $\beta > 1$ and $0 < m < 1/300$, and set $s_n = n^{-1/6+5m}$. There exist constants $c, C > 0$ and $n_0 \in \mathbb{N}$, depending only on β and m, such that*

$$c\, n^{20m} \le K_n(s_n) \le C\, n^{100m} \qquad (n \ge n_0). \tag{20.2}$$

This is [81, Proposition 7.5]. We call it the two-sided bound on positive logarithmic moments; in [81] it is called a positive Laplace window. The normalization of [81] differs from ours in two ways, and neither affects (20.2). First, the Hamiltonian there is βH_n and the partition function carries the factor 2^{-n}, so that the log partition function there is $F_n - n\log 2$. Second, the completion there adds a common centered Gaussian variable D of variance $\beta^2/2$, which has the law of βG. The completed log partition function of [81] therefore has the law of $F_n^c - n\log 2$, the centered logarithmic moment in [81, Proposition 7.5] is K_n, and the tilt $s = n^{-1/6+5m}$, the exponents $20m$ and $100m$, and the range $0 < m < 1/300$ are the same.

We use only two features of (20.2) below: for each fixed small m, the lower bound tends to infinity, and the upper bound grows at most like a fixed small power of n. The particular exponents $20m$ and $100m$ play no further role.

20.2 OpenAI's proof of the two-sided bound

We now summarize the proof of Proposition 20.1 in [81]. The summary is a description of that proof, not a proof, and no other part of this paper depends on it. We follow the notation of [81] where possible, writing $\bar{u}$, $\widehat{\lambda}$, and $\widehat{h}$ for their mass u, strength λ, and field covariance path h. Throughout, $m \in (0, 1/300)$ is fixed and $\delta = m/10000$, as in [81, Section 3.1], and we abbreviate $s = s_n$.

The two-level pressure. Let $H^{(0)}$ and $H^{(1)}$ be independent copies of the completed Hamiltonian H_n^c, that is, independent centered Gaussian processes indexed by Σ_n, each with covariance $nR(x,y)^2/2$. For $0 \le \varepsilon \le \beta^2$, we set

$$\Psi_n(\varepsilon) = \frac{1}{ns}\,\mathbb{E}_{(0)} \log \mathbb{E}_{(1)} \exp\Big\{ s \log \sum_{x \in \Sigma_n} e^{\sqrt{\beta^2-\varepsilon}\,H^{(0)}(x) + \sqrt{\varepsilon}\,H^{(1)}(x)} \Big\}, \tag{20.3}$$

where $\mathbb{E}_{(i)}$ is the expectation over $H^{(i)}$. At $\varepsilon = 0$ the inner expectation is trivial, and $\Psi_n(0) = n^{-1}\mathbb{E}F_n^c$. At $\varepsilon = \beta^2$ the outer expectation is trivial, and $\Psi_n(\beta^2) = (ns)^{-1}\log \mathbb{E}e^{sF_n^c}$. We conclude that

$$K_n(s) = ns\big(\Psi_n(\beta^2) - \Psi_n(0)\big). \tag{20.4}$$

Because $ns = n^{5/6+5m}$, the estimate (20.2) is equivalent to

$$c\, n^{-5/6+15m} \le \Psi_n(\beta^2) - \Psi_n(0) \le C\, n^{-5/6+95m}. \tag{20.5}$$

Gaussian hierarchies. To prove (20.5), OpenAI compares Ψ_n with the pressures of Gaussian hierarchies [81, Section 2.1]. Let $0 = m_0 < m_1 < \cdots < m_{L+1} = 1$, and let b and $\widehat{h}$ be nonnegative nondecreasing functions of the *mass* $\bar{u} \in [0,1)$ that are constant on each interval $[m_j, m_{j+1})$. We call b and $\widehat{h}$ the matrix and the field covariance paths. Their increments at m_j define independent centered Gaussian processes on Σ_n, with covariances

$(b(m_j) - b(m_j-))\, nR(x,y)^2/2$ and $(\widehat{h}(m_j) - \widehat{h}(m_j-))\, nR(x,y)$, where $b(0-) = \widehat{h}(0-) = 0$. Together they form the stage at mass m_j. Let Ξ be the sum of all stages. Starting from $\log(2^{-n}\sum_x e^{\Xi(x)})$, we integrate the stages in the order $j = L, \dots, 0$. The stage at mass m_j is integrated by the Gaussian integration step $X \mapsto m_j^{-1} \log \mathbb{E}_j e^{m_j X}$ of Definition 4.1, where $\mathbb{E}_j$ is the expectation over that stage, and by $X \mapsto \mathbb{E}_j X$ if $m_j = 0$. The pressure $f_n(b, \widehat{h})$ is the resulting number divided by n, with $b(1-)/4$ and $\widehat{h}(1-)/2$ subtracted. It extends to bounded nondecreasing paths by approximation [81, Proposition 2.14]. The associated *path law* is the law of the stages and of a configuration x obtained as follows. Given the earlier stages, the stage at mass m_j is drawn from the tilted Gaussian kernel (4.2) of this recursion; then x is drawn from the Gibbs measure $e^{\Xi(x)} / \sum_y e^{\Xi(y)}$. For $\bar{u} \in [0,1)$, let x^1 and x^2 be two configurations that share the stages up to mass $\bar{u}$ and are conditionally independent given them. We set $S_{\bar{u}} = \mathbb{E}R(x^1, x^2)$, $B_{\bar{u}} = \mathbb{E}R(x^1, x^2)^2$, and $D_{\bar{u}}^2 = B_{\bar{u}} - S_{\bar{u}}^2$ [81, Lemma 2.1]. By [81, Lemma 2.2], a variation of the paths changes f_n by

$$\mathrm{d}f_n = -\frac{1}{4}\int_0^1 B_{\bar{u}}\ \mathrm{d}b(\bar{u})\ \mathrm{d}\bar{u} - \frac{1}{2}\int_0^1 S_{\bar{u}}\ \mathrm{d}\widehat{h}(\bar{u})\ \mathrm{d}\bar{u}.$$

Such hierarchies descend from the Parisi recursion and Guerra's interpolation [59, 91]. The path law and the variance bound with deterministic weights, on which their moment estimates rest, are from [8, Section 2 and Proposition 2.2]. The other basic tools are moment bounds under the path law [81, Lemma 2.3] and a differentiation formula for expectations over finite trees of replicas [81, Proposition 2.5].

In this notation, $\Psi_n(\varepsilon) = f_n(\beta^2 - \varepsilon \mathbf{1}_{[0,s)}, 0) + \beta^2/4 + \log 2$: the matrix covariance $\beta^2 - \varepsilon$ is integrated at mass zero, and the remaining covariance ε at mass s [81, Section 7, opening]. By the variation formula, the derivative of Ψ_n is $\frac{1}{4}\int_0^s B_{\bar{u}}\ \mathrm{d}\bar{u} \in [0, s/4]$ at each point of differentiability, and $0 \le \Psi_n(\varepsilon') - \Psi_n(\varepsilon) \le s(\varepsilon' - \varepsilon)/4$ for $\varepsilon \le \varepsilon'$. The function Ψ_n is nondecreasing.

Comparisons. A comparison [81, Section 3.1, (3)–(6)] has parameters d and E with $n^{-2/3+m} d \lesssim E \le n^{-1/2} d^2$. We write $\underline{x} = \sqrt{n} E/d$; this is the smallest scale at which the profile of its minimizers is examined. The comparison acts on a mass interval $[\bar{u}_0, 1]$, where $\bar{u}_0 = 0$ or $\bar{u}_0 = o(\underline{x})$. It starts from unperturbed paths $(b_0, \widehat{h}_0)$ that are bounded and nondecreasing on $[\bar{u}_0, 1]$ and satisfy $b_0(\bar{u}) = \beta^2 + o(\bar{u}^2)$ and $\widehat{h}_0(\bar{u}) = o(\bar{u}^3)$ for $\bar{u} \gtrsim \underline{x}$ [81, (4)]. Let f be a fixed smooth positive function that equals 1 near 0 and is comparable to $(1 + \log t)/t$ for large t, with logarithmic slope $-tf'(t)/f(t)$ nondecreasing from 0 to 1. We set $e(r) = \widehat{\lambda} f(\widehat{\lambda} r/d^3)$, where $0 \le \widehat{\lambda} \le d^2$ is the *strength*. For a nondecreasing overlap quantile r on $[\bar{u}_0, 1]$, the perturbed paths are $b_0 - e(r)$ and $\widehat{h}_0 + e(r)r$ on $[\bar{u}_0, 1]$, and b_0 and $\widehat{h}_0$ below $\bar{u}_0$; here $e(r)$ is the function $\bar{u} \mapsto e(r(\bar{u}))$. The comparison minimizes

$$\mathcal{F}_{\widehat{\lambda}}(r) = f_n\big(b_0 - e(r), \widehat{h}_0 + e(r)r\big) + \frac{1}{4}\int_{\bar{u}_0}^1 e(r(\bar{u}))\, r(\bar{u})^2\ \mathrm{d}\bar{u}$$

over a compact class of quantiles [81, (5)]. This class is defined by two-sided bounds on the derivative of the inverse function of r and by a prescribed initial value $r(\bar{u}_0) = r_{\min} \geq 0$, which we call the lower radius; $r_{\min} = 0$ when $\bar{u}_0 = 0$. The minimum $\mathcal{M}_{\widehat{\lambda}}$ is attained. The comparison is *admissible* at strength $\widehat{\lambda}$ if the perturbed paths are nonnegative and nondecreasing on $[0, 1)$ for every quantile in the class. For a comparison admissible at every $\widehat{\lambda} \in [0, d^2]$, including $\widehat{\lambda} = 0$, where the perturbed paths are $(b_0, \widehat{h}_0)$, we have $\mathcal{M}_0 = f_n(b_0, \widehat{h}_0)$, and $\mathcal{M}_{\widehat{\lambda}}$ is locally Lipschitz in $\widehat{\lambda}$. At almost every $\widehat{\lambda} > 0$, its derivative in $\log \widehat{\lambda}$ is at most

$$\frac{1}{4} \int_{\bar{u}_0}^{1} e(r(\bar{u})) \left[D_{\bar{u}}^2 + (S_{\bar{u}} - r(\bar{u}))^2 \right] \, \mathrm{d}\bar{u}$$

for every minimizer r at strength $\widehat{\lambda}$, where the moments are those of the perturbed paths [81, Lemma 3.1]. Now let $\widehat{\lambda} \geq n^{-10}$, and suppose that a minimizer at strength $\widehat{\lambda}$ satisfies the *two-sided linear bound* $c\bar{u} \leq r(\bar{u}) \leq C\bar{u}$, with fixed constants $c, C > 0$, for all masses $\bar{u}$ above $A_{\widehat{\lambda}} = n^{\delta} \sqrt{n} E / \sqrt{\widehat{\lambda}}$; this assumption is not needed when $A_{\widehat{\lambda}}$ stays bounded below by a fixed positive constant. The last integral is then at most $n^{10\delta} E$. If the comparison is admissible at every strength and this holds at every strength $\widehat{\lambda} \geq n^{-10}$, integration in $\log \widehat{\lambda}$ gives the comparison bound

$$0 \leq \mathcal{M}_{d^2} - f_n(b_0, \widehat{h}_0) \leq n^{20\delta} E \tag{20.6}$$

[81, Propositions 3.9 and 3.10]. The next two steps establish the two-sided linear bound for the minimizers.

The one-site cavity estimate and the linear profile. Consider a minimizer at a positive strength and a scale $x \to 0$. Suppose that the minimizer satisfies the two-sided linear bound above a fixed multiple of x, and that $x \gtrsim \underline{x}$, $e(x)x^2 \gtrsim nE^2$, and $e(x) \gtrsim x^2$ [81, Section 4, opening]. The last condition says that the strength is comparable to the scale. We write $K = \widehat{h} + bS$, where b and $\widehat{h}$ are the perturbed paths and S is the function $\bar{u} \mapsto S_{\bar{u}}$. Let $\Gamma_K(\bar{u})$ be the second moment, under the path law, of the conditional mean of the spin given the stages up to mass $\bar{u}$, in the hierarchy with $n = 1$, field covariance path K, and no matrix covariance. A one-site cavity estimate [81, Proposition 4.1] shows that a mass $\bar{u}$ with $r(\bar{u}) \leq Cx$, $cx \leq \bar{u} \leq Cx$, and $D_{\bar{u}} = o(x)$ satisfies $S_{\bar{u}} = \Gamma_K(\bar{u}) + o(x^3)$. When only $D_{\bar{u}} = O(x)$ is known, the remainder is $O(x^3)$, uniformly over such masses. The proposition also gives integrated estimates with remainder $o(x^3)$ over these ranges, in mass and in overlap. Its proof removes one spin and controls the resulting susceptibility correction through identities for a fixed number of additional replicas [81, Sections 4.1–4.5]. Using this estimate and the first variation of the comparison, OpenAI proves that, under the same hypotheses, along a subsequence

$$\frac{r(xv)}{x} \longrightarrow \frac{v}{\gamma} \qquad \text{locally uniformly in } v > 0,$$

where γ lies between positive constants that depend only on β and on the bounds on the paths [81, Theorem 5.1]. The proof identifies the limits of the rescaled scalar and overlap distributions [81, Lemma 5.4] and derives a limiting scalar equation [81, Lemmas 5.3 and 5.5]. It then excludes gaps in the limiting support by a first-variation argument [81, Lemma 5.6]. The scalar diffusion identities used there go back to Auffinger and Chen [12, Proposition 3] and to Jagannath and Tobasco [63, Lemma 8.4]. At macroscopic scales, let $c_* > 0$ be fixed. If $x \geq c_*$, $e(x)/x^2 \geq c_*$, and $d^2 \leq \beta^2 - c_*$, then the minimizers converge, in L^1 and locally uniformly on $(0,1)$, to the quantile function of the Parisi measure [81, Proposition 5.9]. This uses the Parisi formula [103], the uniqueness of the Parisi measure [13], and the interval support theorem [69, Theorem 1.1], through [81, Lemma 5.8]. Both statements allow an omitted interval $[0, \bar{u}_0)$ and a positive lower radius $r_{\min}$, provided that the perturbed paths and the competitors used in the first variation are admissible.

Finite propagation. The assumption that the linear bound holds above a multiple of x is removed in [81, Section 6]. There, in addition, $d^2 \leq \beta^2 - c_*$. Either the comparison acts on all masses and b_0 is bounded below by a fixed positive constant, or it acts on $[\bar{u}_0, 1]$, where b_0 is bounded below by a fixed positive constant, and b_0 has a fixed positive jump at $\bar{u}_0$ [81, Definition 6.1]. The selected alternative must retain a fixed positive margin after the initial comparison: a lower bound on the matrix covariance on $[\bar{u}_0, 1]$ and, when the lower interval is omitted, a positive jump in that covariance from $[0, \bar{u}_0)$ to $[\bar{u}_0, 1]$. A comparison on an omitted lower interval may instead be followed by comparisons on all masses if its perturbed matrix path has a global lower bound by a fixed positive constant. Subsequent comparisons on the omitted lower interval have lower radius zero. Suppose that the linear bound fails for some minimizer, at some strength $\widehat{\lambda}$, at some mass at least $A_{\widehat{\lambda}}$. Let x be such a mass within a factor two of the supremum of these masses, so that the bound holds above $2x$. If $e(x)/x^2$ stays bounded below along a subsequence, [81, Theorem 5.1] (when $x \to 0$) or [81, Proposition 5.9] (when x stays bounded below) contradicts the failure, since c and C are chosen outside the bounds of the limiting profiles. Otherwise $e(x)/x^2 \to 0$. In this case OpenAI builds a new comparison, whose unperturbed paths are the perturbed paths of this minimizer, with

$$d' = c_0 x, \qquad E' = \min\{n^{-1/2}(c_0 x)^2,\ n^{-100\delta} e(x) x^3\}, \qquad E' \geq n^{\delta/4} E$$

[81, Lemma 6.2]. Suppose that the new comparison satisfies its comparison bound (20.6), with E' in place of E, and let r' be its minimizer at its final strength. The two minimizers are then close at every scale $y \gtrsim x$ [81, Lemmas 6.3–6.5]. More precisely, for fixed $[t_1, t_2] \subset (0, \infty)$ and $\eta > 0$, with $[t_1 y, t_2 y] \subset (0,1)$ when y is bounded below, the set of masses $\bar{u} \in [t_1 y, t_2 y]$ with $|r(\bar{u}) - r'(\bar{u})| > \eta y$ has Lebesgue measure $o(y)$. It follows that r' satisfies the linear bound above a multiple of x. We then transfer the linear profile of r' at scale x, given by [81, Theorem 5.1] or [81, Proposition 5.9], to r, contradicting the failure. Hence the new comparison violates its comparison bound. By [81, Proposition 3.10], one

of its minimizers then violates the linear bound, and the argument repeats with E' in place of E. The parameter range confines E between fixed multiples of $n^{-5/6+2m}$ and $n^{-1/2}$, and each repetition multiplies it by at least $n^{\delta/4}$. Since $N\delta/4 \geq 5/2 > 1/3 - 2m$ for $N = \lceil 10/\delta \rceil$, at most N repetitions are possible. Let $\kappa_0 > 0$ be a fixed lower bound for the residual matrix covariance and, when required, the residual matrix jump after the initial comparison. Before $n \to \infty$, we fix $c_0 > 0$ so that $Nc_0^2 < \min\{1, \kappa_0/2\}$ and $c_0^2 \leq \beta^2 - c_*$. The total matrix covariance removed by the subsequent comparisons is then less than $\kappa_0/2$, and the total field covariance added is less than 1. Together with the zero lower radius for subsequent comparisons on the omitted lower interval, these bounds preserve nonnegativity, monotonicity, and the fixed bounds on the total covariance throughout the chain [81, Proposition 6.6, proof]. The conclusion is [81, Proposition 6.6]: every minimizer at a positive strength satisfies the linear bound above $A_{\widehat{\lambda}}$, and (20.6) holds when the comparison is admissible at every strength. The linear bound also holds, uniformly in the parameters, for the comparisons with a low-field perturbation used below, which are admissible only at their final strength [81, Corollary 6.7].

The lower bound. Let $\varepsilon = n^{-1/3}$, $d = s$, and $E = n^{-2/3+2m}s$. OpenAI applies a comparison on all masses whose unperturbed paths $(\beta^2 - \varepsilon \mathbf{1}_{[0,s)}, 0)$ are those of $\Psi_n(\varepsilon)$. It combines (20.6) with the chord inequality

$$f_n(b', \widehat{h}') \leq f_n(b, \widehat{h}) + \frac{1}{4} \int_0^1 \frac{(\widehat{h}(\bar{u}) - \widehat{h}'(\bar{u}))^2}{b'(\bar{u}) - b(\bar{u})} \, \mathrm{d}\bar{u},$$

valid for bounded nonnegative nondecreasing paths with $b' \geq b$, $\widehat{h}' = \widehat{h}$ wherever $b' = b$, and a finite right side [81, Lemma 7.1], and with [81, Proposition 6.6], which gives $r \asymp s$ on $[s/2, s]$ at a minimizer. The result is

$$\Psi_n(\varepsilon) - \Psi_n(0) \geq c\, \varepsilon s^3 - n^{20\delta} E = c\, n^{-5/6+15m} - n^{-5/6+7m+20\delta}$$

[81, Section 7.1, (30)]. The second term is smaller than the first by the factor $n^{-8m+20\delta}$. Since Ψ_n is nondecreasing, we have $\Psi_n(\beta^2) - \Psi_n(0) \geq \Psi_n(\varepsilon) - \Psi_n(0)$, and the lower bound in (20.5) follows.

The upper bound. An initial comparison on all masses, with $d_N = n^{-1/6+25m}$, bounds $\Psi_n(\varepsilon_0) - \Psi_n(0)$, where $\varepsilon_0 = n^{-1/3+40m}$ [81, Section 7.2]. Starting at ε_0, OpenAI then iterates

$$\varepsilon \longmapsto \varepsilon + q_\varepsilon, \qquad q_\varepsilon = \tfrac{1}{4} \min\{\varepsilon, \beta^2 - \varepsilon\},$$

until $\beta^2 - \varepsilon \leq n^{-5}$. Below $\beta^2/2$ the parameter ε grows by the factor $5/4$, and above it $\beta^2 - \varepsilon$ shrinks by the factor $3/4$, so $O(\log n)$ steps suffice. Each step uses a comparison on $[s, 1]$ with $d = n^{-1/6+15m}$ and $E = n^{-2/3+10m}d$. Its lower radius is $r_{\min} = \bar{u}_*/(2M(s)) \asymp n^{-1/6-55m-\delta}$, where $\bar{u}_* = (nE)^{-1}$ and $M(\bar{u})$ is the upper bound on the derivative of the

inverse quantile in [81, (5)]. The comparison also includes a low-field perturbation of strength $\theta \in [0,1]$ on the lower interval $[0,s)$, which lowers the matrix covariance on that interval by θc_L and raises the field covariance by $\theta c_L r_0 \varphi$ [81, Section 7.3]. Here $c_L = \frac{1}{2}\min\{\varepsilon, \beta^2 - \varepsilon\}$, $r_0 = d^2 r_{\min}/(2\varepsilon)$, and φ increases from 0 to 1 on $[0,s]$. These comparisons are admissible at their final strength d^2, and their minimizers satisfy the linear bound above the scale d, uniformly in ε and θ [81, Lemma 7.2]. This uses [81, Corollary 6.7] at a fixed positive strength and requires no integration from strength zero. On the lower interval, $S_{\bar{u}} \le C d^3/\varepsilon$ for $0 \le \bar{u} < s$ [81, Lemma 7.3]. Integrating in θ from 0 to 1, and comparing with $\Psi_n(\varepsilon + q_\varepsilon)$ by the chord inequality, OpenAI obtains [81, Lemma 7.4]

$$\Psi_n(\varepsilon + q_\varepsilon) - \Psi_n(\varepsilon) \le n^{20\delta} E + C\Big[\frac{s d^6}{\varepsilon_0} + \frac{d}{n s r_{\min}} (\log n)^3 + n^{-10}\Big].$$

By the bound on the derivative of Ψ_n, the last interval, of length at most n^{-5}, costs at most $s n^{-5}/4$. Summing over the steps, we have [81, Section 7.4]

$$\Psi_n(\beta^2) - \Psi_n(0) \le C\big[n^{-5/6+95m} + n^{-5/6+65m+\delta}(\log n)^4 + n^{-5/6+27m+20\delta}\log n + n^{-5} s\big]. \tag{20.7}$$

Table 20.1 lists the powers of n that enter the two bounds. In (20.7), the first term dominates for large n, because $65m + \delta$ and $27m + 20\delta$ are both smaller than $95m$ when $\delta = m/10000$; the logarithmic factors are absorbed by the remaining powers. Multiplying (20.5) by $ns = n^{5/6+5m}$ turns the exponents $-5/6 + 15m$ and $-5/6 + 95m$ into $20m$ and $100m$, and we obtain (20.2).

Table 20.1. The terms in the proof of (20.5) in [81, Section 7], with logarithmic factors and constants suppressed. Here $s = n^{-1/6+5m}$, $d_N = n^{-1/6+25m}$, $\varepsilon_0 = n^{-1/3+40m}$, $d = n^{-1/6+15m}$, $E = n^{-2/3+10m} d$, and $r_{\min} \asymp n^{-1/6-55m-\delta}$. The middle column refers to the sections of [81].

Term	In [81]	Power of n
main term εs^3 of the lower bound, $\varepsilon = n^{-1/3}$	Section 7.1	$-5/6 + 15m$
error $n^{20\delta} E$ of the lower comparison	Section 7.1	$-5/6 + 7m + 20\delta$
chord term $s\varepsilon_0 d_N^2$ of the initial comparison	Section 7.2	$-5/6 + 95m$
error of the initial comparison	Section 7.2	$-5/6 + 27m + 20\delta$
error $n^{20\delta} E$ of one step	Section 7.4	$-5/6 + 25m + 20\delta$
$s d^6/\varepsilon_0$	Section 7.4	$-5/6 + 55m$
$d/(n s r_{\min})$	Section 7.4	$-5/6 + 65m + \delta$

The order in which the parameters are chosen matters. In [81], m is fixed first and $\delta = m/10000$. Then the constants in the two-sided linear bound, the depth N, and the coefficient c_0 are fixed, and only then does n tend to infinity. The resulting threshold in n

is uniform in ε, θ, the strength, and the minimizer. This uniformity is needed because the number and the locations of the steps depend on n.

20.3 Log-concave augmentation

A two-sided bound on K_n at a single tilt does not by itself control $\mathrm{Var}(F_n)$ or the typical size of $F_n - \mathbb{E}F_n$: a rare large positive deviation can dominate an exponential moment, and a rare negative deviation can dominate the variance without changing the positive exponential moment [81, Section 7.5]. The additional input is that F_n^c has a log-concave law after an independent exponential variable is added. This observation is due to Aronow and Lopatto [8, Proposition 5.1, proof, Steps 1 and 2], and [81, Lemma 7.6] states it in the present setting. Let $\mathcal{E}$ be an exponential random variable with mean one, independent of J and G. We set

$$U_n = F_n^c + \mathcal{E} = F_n + \beta G + \mathcal{E}, \qquad \sigma_n = \bigl(\mathrm{Var}(F_n) + \beta^2/2 + 1\bigr)^{1/2}. \tag{20.8}$$

Lemma 20.2 (Log-concave augmentation)**.** *Let $n \geq 2$ and $\beta \neq 0$. Then U_n has a density $\varrho_n \colon \mathbb{R} \to (0,\infty)$ such that $\log \varrho_n$ is concave. Moreover, $\mathbb{E}U_n = \mathbb{E}F_n + 1$, $\mathrm{Var}(U_n) = \sigma_n^2$, and for $0 \leq t < 1$,*

$$\log \mathbb{E}e^{t(U_n - \mathbb{E}U_n)} = K_n(t) - t - \log(1-t). \tag{20.9}$$

Proof. We follow [8, Proposition 5.1, proof] and [81, Lemma 7.6]. The proof has three steps: we write F_n^c as a convex function of a standard Gaussian vector, use the gauge symmetry of the model to express ϱ_n through a single configuration, and then apply Prékopa's theorem.

Step 1: a Gaussian representation. Let $\bar{n} = n(n-1)/2 + 1$, and let $g = ((J_{ij})_{i<j}, g_0)$ be a standard Gaussian vector in $\mathbb{R}^{\bar{n}}$ with $G = g_0/\sqrt{2}$. For $x \in \Sigma_n$, we define $a_x \in \mathbb{R}^{\bar{n}}$ by $(a_x)_{ij} = \beta x_i x_j/\sqrt{n}$ for $i < j$ and $(a_x)_0 = \beta/\sqrt{2}$. Then $\beta(H_n(x) + G) = a_x \cdot g$, and

$$F_n^c = F(g), \qquad F(g) = \log \sum_{x \in \Sigma_n} e^{a_x \cdot g}.$$

The function F is convex, as the logarithm of a sum of exponentials of linear functions.

Step 2: gauge symmetry. For $y \in \Sigma_n$, let $O_y \colon \mathbb{R}^{\bar{n}} \to \mathbb{R}^{\bar{n}}$ be the linear map with $(O_y g)_{ij} = y_i y_j g_{ij}$ for $i < j$ and $(O_y g)_0 = g_0$. It is diagonal with entries ± 1. It is therefore orthogonal, and it preserves the law of g. Writing $yx = (y_i x_i)_{i \leq n}$, we have $a_x \cdot O_y g = a_{yx} \cdot g$. Summing over x and using that $x \mapsto yx$ is a bijection of Σ_n, we obtain $F(O_y g) = F(g)$.

The density of $U_n = F(g) + \mathcal{E}$ is the convolution of the law of $F(g)$ with the density $e^{-y}\mathbf{1}_{\{y \geq 0\}}$, namely

$$\varrho_n(z) = \mathbb{E}\bigl[e^{F(g) - z}\mathbf{1}_{\{F(g) \leq z\}}\bigr] = e^{-z} \sum_{x \in \Sigma_n} \mathbb{E}\bigl[e^{a_x \cdot g}\mathbf{1}_{\{F(g) \leq z\}}\bigr], \qquad z \in \mathbb{R}.$$

This expectation is finite, since its integrand is at most 1. Fix $x_* \in \Sigma_n$ and write $a_* = a_{x_*}$. For $x \in \Sigma_n$, let $y \in \Sigma_n$ be given by $y_i = x_i x_{*,i}$, so that $yx_* = x$. Using $a_x \cdot g = a_* \cdot O_y g$, $F(g) = F(O_y g)$, and the invariance of the law of g under O_y, we find that

$$\mathbb{E}\big[e^{a_x \cdot g} \mathbf{1}_{\{F(g) \le z\}}\big] = \mathbb{E}\big[e^{a_* \cdot O_y g} \mathbf{1}_{\{F(O_y g) \le z\}}\big] = \mathbb{E}\big[e^{a_* \cdot g} \mathbf{1}_{\{F(g) \le z\}}\big].$$

All 2^n terms of the sum are therefore equal, and

$$\varrho_n(z) = 2^n (2\pi)^{-\bar{n}/2} \int_{\mathbb{R}^{\bar{n}}} \Theta(g, z)\, \mathrm{d}g, \qquad \Theta(g, z) = \exp\Big\{-\frac{|g|^2}{2} + a_* \cdot g - z\Big\} \mathbf{1}_{\{F(g) \le z\}}. \tag{20.10}$$

Step 3: log-concavity and positivity. The set $\{(g, z) \in \mathbb{R}^{\bar{n}+1} : F(g) \le z\}$ is convex, since it is the epigraph of the convex function F. On this set, $\log \Theta(g, z) = -|g|^2/2 + a_* \cdot g - z$ is concave, and Θ vanishes off it; that is, Θ is a log-concave function on $\mathbb{R}^{\bar{n}+1}$. Since $F(g) \ge a_* \cdot g$, we have $a_* \cdot g - z \le 0$ wherever $F(g) \le z$, and hence $0 \le \Theta(g, z) \le e^{-|g|^2/2}$ for every (g, z). Moreover,

$$\int_{\mathbb{R}} \Theta(g, z)\, \mathrm{d}z = e^{-|g|^2/2 + a_* \cdot g - F(g)} \le e^{-|g|^2/2}.$$

Every slice $g \mapsto \Theta(g, z)$ is therefore integrable, and Θ is integrable on $\mathbb{R}^{\bar{n}+1}$ by Tonelli's theorem. By Prékopa's theorem [96, Theorems 6 and 8], the integral of a log-concave function over some of its variables is log-concave in the remaining ones. Then (20.10) implies that $\log \varrho_n$ is concave. To see that $\varrho_n(z) > 0$ for every z, note that $F(g) = \log Z_n + \beta g_0/\sqrt{2}$. For fixed J, the set of g_0 with $F(g) \le z$ is a half-line, since $\beta \ne 0$. By Fubini's theorem, $\{g : F(g) \le z\}$ has positive Lebesgue measure, and the integrand in (20.10) is positive on it.

Finally, J, G, and $\mathcal{E}$ are independent, $\mathbb{E}G = 0$, $\mathbb{E}\mathcal{E} = \mathrm{Var}(\mathcal{E}) = 1$, and $\mathrm{Var}(\beta G) = \beta^2/2$. It follows that $\mathbb{E}U_n = \mathbb{E}F_n + 1$ and $\mathrm{Var}(U_n) = \sigma_n^2$. For $0 \le t < 1$, we have $\mathbb{E}e^{t(\mathcal{E}-1)} = e^{-t}/(1-t)$, and (20.9) follows from the independence of F_n^c and $\mathcal{E}$. □

The next lemma collects the estimates for one-dimensional log-concave laws that we need. It is a variant of [81, Lemma 7.7]; the tail bound in its proof is also in [8, Proposition 5.1, proof, Step 3].

Lemma 20.3 (One-dimensional log-concave bounds)**.** *There exist universal constants $c_1, c_2, p_*, C_1 > 0$ with the following property. Let Z be a real random variable with mean μ and finite variance $\sigma^2 > 0$, and suppose that Z has a density $\varrho\colon \mathbb{R} \to (0, \infty)$ such that $\log \varrho$ is concave. Then*

$$\log \mathbb{E}e^{t(Z-\mu)} \le \log 3 \qquad \Big(|t| \le \frac{c_1}{2\sigma}\Big), \tag{20.11}$$

$$\log \mathbb{E}e^{t(Z-\mu)} \ge t c_2 \sigma + \log p_* \qquad (t \ge 0), \tag{20.12}$$

$$\mathbb{P}\{z - r \le Z \le z + r\} \le \frac{2C_1 r}{\sigma} \qquad (z \in \mathbb{R},\ r \ge 0). \tag{20.13}$$

In (20.12)*, the left side may be* $+\infty$*.*

Proof. We show that the lemma holds with

$$c_1 = \frac{\log 3}{4}, \qquad C_2 = \frac{48}{c_1^4}, \qquad c_2 = \frac{1}{4\sqrt{C_2}}, \qquad p_* = c_2^2,$$

and

$$C_1 = \Big(16 + \frac{4}{c_3^3}\Big)^{1/2}, \qquad c_3 = \frac{\log 2}{4}.$$

We first prove a tail bound and then the three assertions in order.

Step 1: an exponential tail bound. The function $(x, y) \mapsto \mathbf{1}_{\{y \ge x\}} \varrho(y)$ is log-concave on $\mathbb{R}^2$. By Prékopa's theorem [96, Theorems 6 and 8], the survival function $Q(x) = \mathbb{P}(Z \ge x)$ is log-concave. It is positive, because $\varrho > 0$. By Chebyshev's inequality,

$$Q(\mu - 2\sigma) \ge 1 - \mathbb{P}(|Z - \mu| \ge 2\sigma) \ge \frac{3}{4}, \qquad Q(\mu + 2\sigma) \le \mathbb{P}(|Z - \mu| \ge 2\sigma) \le \frac{1}{4}.$$

Let $x \ge \mu + 2\sigma$. Since $\log Q$ is concave, the slope of its secant over $[\mu + 2\sigma, x]$ is at most its slope over $[\mu - 2\sigma, \mu + 2\sigma]$, which is at most $(\log \frac{1}{4} - \log \frac{3}{4})/(4\sigma) = -c_1/\sigma$. This gives

$$Q(x) \le \frac{1}{4} e^{-c_1(x - \mu - 2\sigma)/\sigma} = \frac{\sqrt{3}}{4} e^{-c_1(x-\mu)/\sigma},$$

where we used $e^{2c_1} = \sqrt{3}$. The variable $-Z$ satisfies the same hypotheses, with mean $-\mu$. Applying the same bound to it and adding, we find $\mathbb{P}(|Z - \mu| \ge u) \le (\sqrt{3}/2) e^{-c_1 u/\sigma}$ for $u \ge 2\sigma$. For $0 \le u < 2\sigma$, we have $2e^{-c_1 u/\sigma} \ge 2/\sqrt{3} > 1$. We conclude that

$$\mathbb{P}(|Z - \mu| \ge u) \le 2e^{-c_1 u/\sigma} \qquad (u \ge 0). \tag{20.14}$$

Step 2: small tilts. Let $|t| \le c_1/(2\sigma)$ and $\ell = |t|$. Writing $e^{\ell|Z-\mu|} = 1 + \int_0^\infty \ell e^{\ell u} \mathbf{1}_{\{|Z-\mu| > u\}} \, \mathrm{d}u$ and using (20.14), we obtain

$$\mathbb{E} e^{t(Z-\mu)} \le \mathbb{E} e^{\ell|Z-\mu|} \le 1 + 2\ell \int_0^\infty e^{(\ell - c_1/\sigma)u} \, \mathrm{d}u = 1 + \frac{2\ell}{c_1/\sigma - \ell} \le 3.$$

In the last inequality, we used $c_1/\sigma - \ell \ge c_1/(2\sigma) \ge \ell$. This proves (20.11).

Step 3: positive tilts. We use the fourth-moment and positive-part argument of [81, Lemma 7.7, proof]. Let $\hat{Z} = (Z - \mu)/\sigma$, so that $\mathbb{E}\hat{Z} = 0$ and $\mathbb{E}\hat{Z}^2 = 1$. By (20.14),

$$\mathbb{E}\hat{Z}^4 = \int_0^\infty 4u^3 \, \mathbb{P}(|\hat{Z}| \ge u) \, \mathrm{d}u \le 8 \int_0^\infty u^3 e^{-c_1 u} \, \mathrm{d}u = \frac{48}{c_1^4} = C_2.$$

By Hölder's inequality with exponents $3/2$ and 3, we have $1 = \mathbb{E}\big[|\hat{Z}|^{2/3} |\hat{Z}|^{4/3}\big] \le (\mathbb{E}|\hat{Z}|)^{2/3} (\mathbb{E}\hat{Z}^4)^{1/3}$, which implies that $\mathbb{E}|\hat{Z}| \ge C_2^{-1/2}$. The positive part $\hat{Z}_+$ satisfies

$\mathbb{E}\hat{Z}_+ = \mathbb{E}|\hat{Z}|/2 \geq C_2^{-1/2}/2 = 2c_2$, because $\mathbb{E}\hat{Z} = 0$. We have $\hat{Z}_+ < c_2$ on $\{\hat{Z} < c_2\}$. Using this, the Cauchy–Schwarz inequality, and $\mathbb{E}\hat{Z}_+^2 \leq 1$, we obtain

$$2c_2 \leq \mathbb{E}\hat{Z}_+ \leq c_2 + \mathbb{E}\big[\hat{Z}_+ \mathbf{1}_{\{\hat{Z} \geq c_2\}}\big] \leq c_2 + \mathbb{P}(\hat{Z} \geq c_2)^{1/2}.$$

Rearranging, we find $\mathbb{P}(\hat{Z} \geq c_2) \geq c_2^2 = p_*$. For $t \geq 0$, restricting the expectation to $\{\hat{Z} \geq c_2\}$, we obtain $\mathbb{E}e^{t(Z-\mu)} = \mathbb{E}e^{t\sigma\hat{Z}} \geq e^{tc_2\sigma}p_*$, which is (20.12).

Step 4: the density. We adapt the log-density secant argument of [81, Lemma 7.7, proof], using an arbitrary point in place of a mode. Fix $z \in \mathbb{R}$, and set $\ell = 4/\varrho(z)$. Suppose that $\varrho(z+\ell) \geq \varrho(z)/2$. Since $\log\varrho$ is concave, for $0 \leq y \leq \ell$ we have

$$\log\varrho(z+y) \geq \Big(1 - \frac{y}{\ell}\Big)\log\varrho(z) + \frac{y}{\ell}\log\varrho(z+\ell) \geq \log\frac{\varrho(z)}{2},$$

so that $\int_z^{z+\ell}\varrho \geq \ell\varrho(z)/2 = 2$, which is impossible. We conclude that $\varrho(z+\ell) < \varrho(z)/2$. For $y \geq \ell$, the slope of the secant of $\log\varrho$ over $[z, z+y]$ is at most its slope over $[z, z+\ell]$, which is less than $-(\log 2)/\ell = -c_3\varrho(z)$. Exponentiating, we find that $\varrho(z+y) \leq \varrho(z)e^{-c_3\varrho(z)y}$ for $y \geq \ell$, and by the same argument applied to $y \mapsto \varrho(z-y)$, also $\varrho(z-y) \leq \varrho(z)e^{-c_3\varrho(z)y}$ for $y \geq \ell$. Since the mean minimizes the mean square distance, we have

$$\begin{aligned}\sigma^2 \leq \mathbb{E}(Z-z)^2 &\leq \ell^2\,\mathbb{P}(|Z-z| < \ell) + 2\int_\ell^\infty y^2\varrho(z)e^{-c_3\varrho(z)y}\,\mathrm{d}y \\ &\leq \frac{16}{\varrho(z)^2} + \frac{4}{c_3^3\varrho(z)^2} = \frac{C_1^2}{\varrho(z)^2}.\end{aligned}$$

In the second inequality, we bounded $(Z-z)^2$ by ℓ^2 on $\{|Z-z| < \ell\}$ and used the two exponential bounds on the complement. This shows that $\varrho(z) \leq C_1/\sigma$ for every $z \in \mathbb{R}$, and integrating over $[z-r, z+r]$ proves (20.13). □

20.4 Proof of the fluctuation theorem

The argument below is that of [81, Section 7.6], with two elementary changes: we use the constant bound (20.11) for small tilts in place of the quadratic bound of [81, Lemma 7.7], and we remove the added noise through the inequality (20.17) below.

Proof of Theorem 1.6 from Proposition 20.1. The proof has three steps. We first bound σ_n above and below, using the two halves of (20.2) together with Lemmas 20.2 and 20.3. We then subtract the variance of the added noise to obtain the bounds on $\mathrm{Var}(F_n)$. Finally, we use the bound (20.13) on the density of U_n to show that $|F_n - \mathbb{E}F_n|$ is rarely smaller

than $n^{1/6-\eta}$. Throughout, c_1, c_2, p_*, C_1 are the universal constants of Lemma 20.3, and we write

$$d_\beta = \frac{\beta^2}{2} + 1, \qquad \text{so that} \qquad \operatorname{Var}(F_n) = \sigma_n^2 - d_\beta.$$

By Lemma 20.2, the variable $Z = U_n$ satisfies the hypotheses of Lemma 20.3, with $\mu = \mathbb{E}U_n$ and $\sigma = \sigma_n \geq 1$.

Step 1: bounds on σ_n. Fix $0 < m < 1/300$, and let c, C, n_0 be as in Proposition 20.1. Let $\omega(t) = -t - \log(1-t)$ for $0 \leq t < 1$. Since $\omega(t) = \sum_{k\geq 2} t^k/k$, we have

$$0 \leq \omega(t) \leq \frac{t^2}{2(1-t)} \leq t^2 \qquad \Big(0 \leq t \leq \frac{1}{2}\Big).$$

By (20.9), $\log \mathbb{E}e^{t(U_n - \mathbb{E}U_n)} = K_n(t) + \omega(t)$. Let $n_1 \geq n_0$ be such that $s_n \leq 1/2$ and $c\, n^{20m} > \log 3$ for all $n \geq n_1$, and let $n \geq n_1$. If $s_n \leq c_1/(2\sigma_n)$, then (20.11) and $\omega \geq 0$ imply that $K_n(s_n) \leq \log 3$, which contradicts the lower bound in (20.2). It follows that $\sigma_n > c_1/(2s_n)$. On the other hand, (20.12) and the upper bound in (20.2) give

$$s_n c_2 \sigma_n + \log p_* \leq K_n(s_n) + \omega(s_n) \leq C\, n^{100m} + 1.$$

Using $p_* < 1$ and $n^{100m} \geq 1$, we find $\sigma_n \leq C' n^{100m}/s_n$ with $C' = (C + 1 + \log(1/p_*))/c_2$. We have shown that

$$\frac{c_1}{2}\, n^{1/6-5m} \leq \sigma_n \leq C' n^{1/6+95m} \qquad (n \geq n_1). \tag{20.15}$$

Now fix $\eta > 0$, and let $m = \min\{1/600, \eta/200\}$. Then $0 < m < 1/300$, $5m < \eta$, and $95m < \eta$. Since $n^{\eta-5m}$ and $n^{\eta-95m}$ tend to infinity, (20.15) implies that there exists $n_2(\beta, \eta)$ such that

$$n^{1/6-\eta} \leq \sigma_n \leq n^{1/6+\eta} \qquad (n \geq n_2(\beta, \eta)). \tag{20.16}$$

Here m depends only on η and is chosen before $n \to \infty$; the threshold n_2 depends only on β and η.

Step 2: the standard deviation and the variance. Fix $\eta > 0$. By (20.16) and $\operatorname{Var}(F_n) \leq \sigma_n^2$, we have $\operatorname{Var}(F_n)^{1/2} \leq n^{1/6+\eta}$ for $n \geq n_2(\beta, \eta)$. For the lower bound, let $\eta' = \min\{\eta/2, 1/12\}$. By (20.16) with η' in place of η, we have $\sigma_n^2 \geq n^{1/3-2\eta'}$ for $n \geq n_2(\beta, \eta')$. Since $1/3 - 2\eta' \geq 1/6 > 0$, we have $d_\beta \leq n^{1/3-2\eta'}/2$ for all sufficiently large n, and then

$$\operatorname{Var}(F_n) = \sigma_n^2 - d_\beta \geq \frac{1}{2} n^{1/3-2\eta'} \geq n^{1/3-2\eta}.$$

The last inequality holds once $n^{2\eta-2\eta'} \geq 2$, which is true eventually because $2\eta - 2\eta' \geq \eta > 0$. Taking square roots, we conclude that $n^{1/6-\eta} \leq \operatorname{Var}(F_n)^{1/2} \leq n^{1/6+\eta}$ for all sufficiently large n, for every fixed $\eta > 0$. This is (1.4). Applying it with $\eta/2$ in place of η and squaring, we obtain $n^{1/3-\eta} \leq \operatorname{Var}(F_n) \leq n^{1/3+\eta}$ for all sufficiently large n, which is (1.5).

Step 3: the typical magnitude. We have $U_n - \mathbb{E}U_n = (F_n - \mathbb{E}F_n) + \beta G + (\mathcal{E} - 1)$. Let $r \geq 0$ and $v > 0$. On the event $\{|F_n - \mathbb{E}F_n| \leq r\} \cap \{|\beta G| \leq v\} \cap \{|\mathcal{E} - 1| \leq v\}$, we have $|U_n - \mathbb{E}U_n| \leq r + 2v$. A union bound, (20.13) with $z = \mathbb{E}U_n$, and Chebyshev's inequality for βG and $\mathcal{E}$, whose variances are $\beta^2/2$ and 1, give

$$\mathbb{P}(|F_n - \mathbb{E}F_n| \leq r) \leq \frac{2C_1(r + 2v)}{\sigma_n} + \frac{d_\beta}{v^2}. \tag{20.17}$$

Fix $\eta > 0$, and set $\eta_1 = \min\{\eta/4, 1/48\}$, $r = n^{1/6-\eta}$, and $v = n^{\eta_1}$. By (20.16) with η_1 in place of η, we have $\sigma_n \geq n^{1/6-\eta_1}$ for $n \geq n_2(\beta, \eta_1)$, and then (20.17) gives

$$\mathbb{P}\big(|F_n - \mathbb{E}F_n| \leq n^{1/6-\eta}\big) \leq 2C_1\big(n^{\eta_1-\eta} + 2n^{2\eta_1-1/6}\big) + d_\beta n^{-2\eta_1}.$$

Since $\eta_1 - \eta \leq -3\eta/4$ and $2\eta_1 - 1/6 \leq -1/8$, the right side tends to zero. For the upper tail, Chebyshev's inequality and the bound on the variance from Step 2 give

$$\mathbb{P}\big(|F_n - \mathbb{E}F_n| > n^{1/6+\eta}\big) \leq \frac{\mathrm{Var}(F_n)}{n^{1/3+2\eta}} \leq n^{-\eta}$$

for all sufficiently large n. Combining the two bounds, we find that $\mathbb{P}\{n^{1/6-\eta} \leq |F_n - \mathbb{E}F_n| \leq n^{1/6+\eta}\} \to 1$ for every fixed $\eta > 0$, which is (1.6). This completes the proof. □

SECTION 21

The finite-size mean bias

In this section, we prove Corollary 1.7, the lower bound with exponent $5/6$ on the mean bias $\Delta_{n,\beta} = P_\beta - p_n(\beta)$ stated in the introduction. Combined with the upper bound of Aronow and Lopatto (Theorem 4.11), it shows that, for every fixed $\beta > 1$ and $\varepsilon > 0$, we have $c_{\beta,\varepsilon} n^{-5/6-\varepsilon} \le \Delta_{n,\beta} \le C_b n^{-7/10}$ for all sufficiently large n. If $\Delta_{n,\beta} = n^{-a+o(1)}$ for some exponent a, then $7/10 \le a \le 5/6$ (Remark 21.3). Corollary 1.7 follows from the lower bound on the variance in Theorem 1.6 and from a variance–bias inequality of Aronow and Lopatto (Proposition 21.1). The deduction itself is not new: Aronow and Lopatto record it in [6, Section 1.6, p. 11], as a consequence of the variance exponent stated in [81, Theorem 1.1]. Here we formally verify this deduction with a Lean proof that follows the argument of Section 21.2 and composes the formal proof of Theorem 1.6 with a formal proof of Proposition 21.1 (see Section 22.2). We include the short written proof for completeness, and in Remarks 21.2 and 21.3 we discuss what it does not determine. Section 23 returns to the exact exponent of the mean bias.

Recall from Section 3 that $p_n(\beta) = n^{-1}\mathbb{E}F_n$, with $F_n = \log Z_n$, that $P_\beta = \lim_n p_n(\beta)$, and that $\Delta_{n,\beta} = P_\beta - p_n(\beta) \ge 0$ for every $n \ge 2$ (see (3.4)).

21.1 The variance–bias inequality

Proposition 21.1 (Variance–bias inequality)**.** *Fix $\beta > 1$. There exists a constant $c_\beta > 0$ such that, for every $n \ge 2$,*

$$\Delta_{n,\beta} \ge c_\beta \Big(\frac{\mathrm{Var}(F_n) + \beta^2/2}{n} \Big)^{5/4}. \tag{21.1}$$

This is [8, Remark 7.14, (7.68)], which restates [8, Lemma 7.4] in the normalization of the log partition function. The normalization of [8, (1.1)] is ours: the Hamiltonian is $n^{-1/2}\sum_{i<j} J_{ij}x_i x_j$, the partition function is a sum over Σ_n with counting measure, β multiplies the Hamiltonian, and the Parisi value has the terminal condition $\log(2\cosh z)$. The inequality (21.1) is [8, (7.68)] verbatim; there it is stated for every $n \ge 1$. The term $\beta^2/2$ is the variance of the completion βG (Lemma 3.1), and $\mathrm{Var}(F_n) + \beta^2/2 = \mathrm{Var}(F_n^c)$ for the completed log partition function $F_n^c = F_n + \beta G$.

We recall the argument of [8, Lemma 7.4], which works with the variable F_n^c/β. For $t > 0$, let $\Lambda(t) = (nt)^{-1}\log \mathbb{E}e^{tF_n^c/\beta}$, and let $\Lambda(0) = \mathbb{E}F_n^c/(\beta n) = p_n(\beta)/\beta$. Since F_n^c/β is a convex Lipschitz function of a standard Gaussian vector, [8, Lemma 5.2],

which rests on W.-K. Chen's convexity theorem for Gaussian logarithmic moments [37, Theorem 1], shows that Λ is convex with $\Lambda'(0) = V/2$, where $V = \mathrm{Var}(F_n^c)/(\beta^2 n)$. By convexity, $\Lambda(t) \geq \Lambda(0) + tV/2$: a gain linear in t and proportional to the variance. On the other hand, the finite-size comparison [8, (7.1)] and the fifth-order bound for a scalar trial profile [8, Lemma 7.2, (7.3)] give $\Lambda(t) \leq P_\beta/\beta + Ct^5$ for small $t > 0$. Here C and the range $0 < t \leq t_*$ of the trial bound depend only on β. Subtracting, we see that $\Delta_{n,\beta}/\beta$ is at least $tV/2 - Ct^5$. The gradient of F_n^c/β is a convex combination of vectors of squared norm $n/2$, and the Gaussian Poincaré inequality [28, Theorem 3.20], which is the Gaussian case of the Brascamp–Lieb bound in Proposition 4.21 [30], gives $\mathrm{Var}(F_n^c) \leq \beta^2 n/2$; also $\mathrm{Var}(F_n^c) \geq \mathrm{Var}(\beta G) = \beta^2/2$. We conclude that $0 < V \leq 1/2$. With $c' = \min\{1, t_*, (4C)^{-1/4}\}$ and $t = c'V^{1/4}$, we have $t \leq t_*$ and

$$\frac{\Delta_{n,\beta}}{\beta} \geq \frac{c'}{2} V^{5/4} - C(c')^5 V^{5/4} \geq \frac{c'}{4} V^{5/4}.$$

Since $V = \beta^{-2}(\mathrm{Var}(F_n) + \beta^2/2)/n$, multiplying by β gives (21.1) with $c_\beta = (c'/4)\beta^{-3/2}$, independent of n.

21.2 Proof of the mean-bias bound

Proof of Corollary 1.7. Fix $\beta > 1$ and $\varepsilon > 0$. By the lower bound in (1.5), applied with $\eta = 4\varepsilon/5$, there exists $n_0(\beta, \varepsilon) \geq 2$ such that

$$\mathrm{Var}(F_n) \geq n^{1/3 - 4\varepsilon/5} \qquad (n \geq n_0(\beta, \varepsilon)).$$

Since $t \mapsto t^{5/4}$ is increasing on $[0, \infty)$, we may drop the term $\beta^2/2$ in (21.1). Inserting the variance bound, we obtain, for $n \geq n_0(\beta, \varepsilon)$,

$$\Delta_{n,\beta} \geq c_\beta \big(n^{1/3 - 4\varepsilon/5 - 1}\big)^{5/4} = c_\beta\, n^{-5/6 - \varepsilon},$$

where we used $\frac{5}{4}\big(-\frac{2}{3} - \frac{4\varepsilon}{5}\big) = -\frac{5}{6} - \varepsilon$. Multiplying by n, we find $nP_\beta - \mathbb{E}F_n \geq c_\beta n^{1/6 - \varepsilon}$. This proves Corollary 1.7 with $c_{\beta,\varepsilon} = c_\beta$ and this threshold $n_0(\beta, \varepsilon)$; only the threshold depends on ε. □

If we use instead the earlier lower bound $\mathrm{Var}(F_n) \geq c\, n^{4/15}$ of [8, Theorem 1.4], the same computation yields $\Delta_{n,\beta} \geq c\, n^{-11/12}$, which is how the lower bound on the mean bias in [8, Theorem 1.4] is obtained there. Using the variance exponent $1/3$ gives $5/6 + \varepsilon$ in place of $11/12$, as already observed in [6, Section 1.6, p. 11].

21.3 What the bound does not determine

Remark 21.2 (A positive variance prefactor). If there were a constant $k_\beta > 0$ such that $\mathrm{Var}(F_n) \geq k_\beta n^{1/3}$ for all sufficiently large n, the same computation would give

$$\Delta_{n,\beta} \geq c_\beta k_\beta^{5/4}\, n^{-5/6} \tag{21.2}$$

for all sufficiently large n. This is [6, (1.4), p. 11], where Aronow and Lopatto note that it would follow from the positive limit of $n^{-1/3}\,\mathrm{Var}(F_n)$ claimed in [82, Theorem 1.2]. Theorem 1.6 does not give such a prefactor: a variance equal to $n^{1/3}/\log n$, for example, is of the form $n^{1/3+o(1)}$, and inserting it in (21.1), we obtain (21.2) only up to the factor $(\log n)^{-5/4}$. We therefore prove the bound with exponent $5/6 + \varepsilon$, not (21.2).

Remark 21.3 (The exponent of the mean bias). For every fixed $\beta > 1$, Aronow and Lopatto prove that $\Delta_{n,\beta} \leq C_b n^{-7/10}$ for $n \geq 2$ [6, Theorem 1.1]. Combined with Corollary 1.7, this bound implies that if $\Delta_{n,\beta} = n^{-a+o(1)}$ for some exponent a, then $7/10 \leq a \leq 5/6$ [6, Section 1.6, p. 11]. The exact exponent is not determined by these bounds. Franz, Parisi, and Ricci-Tersenghi predict $a = 3/4$, possibly with logarithmic corrections [56, (13)–(17)]; this prediction is consistent with both bounds.

PART VI

Formal verification and open problems

SECTION 22

Formal verification

The assembled Palomar submission `https://github.com/pmaronow/fluctuations-palomar` exposes five claims through `Solution.lean`, selected by `comparator.json`; its independent `Challenge.lean` contains those five claims. The combined package passed the recorded checks on 19 October 2026 (20 October UTC), for proof-source commit `5551e62264a270fd5dcd396f36389eaf77eddebc`. The default build succeeded using unchanged compiled objects, and a strict axiom audit accepted all five exports with only `propext`, `Quot.sound`, and `Classical.choice`. A fresh protected Comparator run compiled the independent Challenge, compared the five statements and their complete reachable helper dependencies, and replayed the Solution export through Lean's kernel, NanoDa, and con-ron. All three accepted it. The records are published at `https://github.com/pmaronow/fluctuations-palomar/tree/351cc102439ff1d741d240391dfb62258605b4d2/verification`. All Lean sources and build/comparison controls are unchanged from the verified proof-source commit. These checks concern the combined five-claim package; the source-package checks below, including the historical 47-claim upper-bound check, retain their original scope.

In this section, we describe formal proofs of the four results Theorems 1.1 and 1.6 and Corollaries 1.2 and 1.7 of the lower bound and the fluctuation part in the Lean 4 proof assistant [48], built on the Mathlib library [108]. No argument elsewhere in the paper depends on this section. In Section 22.1, we recall what OpenAI's public formalization proves. In Section 22.2, we state the four formal theorems, explain how they correspond to the statements of this paper, and describe how they were checked. Section 22.3 compares the formal proof of Theorem 1.1 with the proof in this paper, and Section 22.4 describes the mathematics that the formal proof of Theorem 1.6 had to supply beyond OpenAI's formalization. Section 22.5 describes the upper-bound formalization, and Section 22.6 lists what is not formally verified.

22.1 OpenAI's formalization

OpenAI's public repository [83] contains Lean formalizations accompanying [77, 81]. For the barrier, the theorem `OAI.SK.main`, in the file `lean/OAI/Probability/SKBarriers/Main.lean`, states that for every $\beta > 1$ two quantities tend to one: the Gibbs mass, averaged over the disorder, of the initial states whose total variation distance from

π^J at continuous time $\exp(n^{1/10000})$ is greater than $1/4$, and the same mass with the distance after $\lfloor \exp(n^{1/10000}) \rfloor$ discrete attempts. This is the formal counterpart of [77, Theorem 1.1]; the repository's scope document `lean/docs/227.md` describes it.

For the fluctuations, the theorem `main_of_positiveLaplaceWindow`, at line 294 of `lean/OAI/Probability/SKFluctuations/EventuallyNatPos.lean`, has the hypothesis `hw:PositiveLaplaceWindow` and the conclusion `MainFluctuationStatement`. Both propositions are defined in the file `Spin.lean` of the same directory, at lines 453 and 531. The proposition `PositiveLaplaceWindow` is the statement of Proposition 20.1, for every $\beta > 1$ and every $0 < m < 1/300$, and `MainFluctuationStatement` is the conjunction of the three assertions of Theorem 1.6, for every $\beta > 1$ and every $\eta > 0$. `MainFluctuationStatement` uses the log partition function before completion, normalized as $\log(2^{-n} \sum_x e^{\beta H_n(x)}) = F_n - n \log 2$, which has the same variance and the same centered fluctuations as F_n. `PositiveLaplaceWindow` uses its Gaussian completion $F_n^c - n \log 2$, whose logarithmic moment generating function after centering is K_n. OpenAI's formalization proves the deduction of Theorem 1.6 from Proposition 20.1 given in [81, Sections 7.5 and 7.6], of which Sections 20.3 and 20.4 give a variant, but not Proposition 20.1 itself. Aronow and Lopatto inspected this directory and found no proof of the hypothesis there [6, Section 1.6, footnote 1]. In the paper [81], by contrast, Proposition 20.1 is proved, as summarized in Section 20.2.

22.2 Statements of the formal results

We originally formalized the four results in two source packages, one for Theorem 1.1 and Corollary 1.2 and one for Theorem 1.6 and Corollary 1.7. Both use Lean version `v4.35.0-rc2` and Mathlib at commit `065356127b1dc0016f66b7283ce0ce2c4055aa55`. Both include OpenAI's Lean sources at commit `fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb` as a dependency, adapted to the module system of this version of Lean. The second package also depends on the first, which contains the formal proof of the variance–bias inequality.

The barrier theorem. The theorem `SKBarriers2.mainStatementSevenThirty` has type `OAI.SK.Strengthened.MainStatementAt(7/30)`. This proposition states that for every $\beta > 1$ there exists $A > 0$ with the following property: for every $\tau \in (0, 1/2)$, the two quantities

$$\int \sum_{x \in \Sigma_n} \pi^J(x)\, \mathbf{1}_{\{d^J(x, \exp(An^{7/30})) > \tau\}}\, \mathbb{P}(\mathrm{d}J), \qquad \int \sum_{x \in \Sigma_n} \pi^J(x)\, \mathbf{1}_{\{d^J_{\mathrm{disc}}(x, \lfloor \exp(An^{7/30}) \rfloor) > \tau\}}\, \mathbb{P}(\mathrm{d}J)$$

tend to one as $n \to \infty$. These are the two probabilities in Theorem 1.1, written as in (3.16), and A plays the role of A_β. As in Theorem 1.1, the coefficient is chosen before the threshold τ, and the same coefficient serves both clocks. The disorder law,

the Hamiltonian, the Gibbs measure, the two heat-bath kernels, and the total variation distances are those of OpenAI's development. They agree with the definitions in Section 3: the partition function Z_n is a sum over Σ_n with counting measure, one discrete step is an update attempt at a uniformly chosen site, an attempt that leaves the configuration unchanged is counted, and the continuous-time kernel is given by (3.15). Both displayed quantities are instances of a definition in the development that takes the coefficient, the exponent of n, and the threshold as arguments. The development proves that the instance with coefficient 1, exponent $1/10000$, and threshold $1/4$ is the corresponding quantity of OpenAI's theorem. The theorem `SKBarriers2.subcriticalStatementSevenThirty`, of type `OAI.SK.Strengthened.SubcriticalStatementBelow(7/30)`, is Corollary 1.2: the same conclusion with coefficient one, at the time horizons $\exp(n^\kappa)$ and $\lfloor \exp(n^\kappa) \rfloor$, for every $0 < \kappa < 7/30$. The two theorems have no hypotheses.

The fluctuation theorem. The fluctuation theorem is the declaration `OAI.SKFluctuations.LaplaceWindowFormalization.main_fluctuation_statement_unconditional_original_q` in the module `SKProof.OriginalQUnconditionalMain`. Its type is `OAI.SKFluctuations.MainFluctuationStatement`, the definition of OpenAI's development described in Section 22.1, and it has no hypotheses. The identifier is only a locator; its parts are not used as terminology in this paper. In `MainFluctuationStatement`, the bounds on the standard deviation and on the variance come with separate thresholds in n, and their maximum serves as the threshold $n_0(\beta, \eta)$ of Theorem 1.6. The proof of the theorem establishes `PositiveLaplaceWindow` and then applies OpenAI's theorem `main_of_positiveLaplaceWindow`. In other words, our formal proof of Theorem 1.6 consists of a formal proof of Proposition 20.1 together with OpenAI's formal proof of the deduction.

The mean-bias corollary. The declaration `OAI.SKFluctuations.LaplaceWindowFormalization.mean_bias_lower_bound_unconditional`, in the module `SKProof.MeanBiasLowerBound`, has type `OAI.SKFluctuations.MeanBias.LowerBoundStatement`. This proposition states that for every $\beta > 1$ and every $\varepsilon > 0$, there exist $c > 0$ and n_0 such that

$$c\, n^{-5/6-\varepsilon} \le \lim_{m\to\infty} p_m(\beta) - p_n(\beta) \qquad (n \ge n_0),$$

where we write $p_n(\beta) = n^{-1}\mathbb{E}\log\sum_{x\in\Sigma_n} e^{\beta H_n(x)}$, with OpenAI's disorder law and Hamiltonian, which are those of Section 3. Since $P_\beta = \lim_m p_m(\beta)$ by definition, this is Corollary 1.7. The development proves that this limit exists and equals the infimum of the Parisi functional, as in Theorem 4.4.

To prove this bound, we follow the proof of Corollary 1.7 in Section 21.2. We combine our formal proof of Theorem 1.6 with the theorem `QuantitativePaper.short_variance_mean` of the Lean development [7] accompanying [6], which is a formal version of

Proposition 21.1. That theorem concerns the variable F_n/β: for every $\beta_0 > 1$, there exists $c > 0$ such that

$$c\Big(\frac{\mathrm{Var}(F_n/\beta) + 1/2}{n}\Big)^{5/4} \le \frac{\Delta_{n,\beta}}{\beta}$$

for every $n \ge 1$ and every $\beta \ge \beta_0$. Since $\mathrm{Var}(F_n/\beta) + 1/2 = \beta^{-2}(\mathrm{Var}(F_n) + \beta^2/2)$, at a fixed β this is (21.1) with $c_\beta = c\beta^{-3/2}$. The module `SKProof.MeanBiasBridge` proves the identities that connect the two developments: their disorder laws and Hamiltonians coincide, their expected pressures agree, and OpenAI's log partition function is $F_n - n\log 2$, so that its variance is $\mathrm{Var}(F_n)$. As in Section 21.2, the lower bound on the variance in Theorem 1.6, with $\eta = 4\varepsilon/5$, then gives the exponent $-5/6 - \varepsilon$.

Axioms and the independent check. The four theorems use only the three standard axioms `propext`, `Classical.choice`, and `Quot.sound` of Lean's logic. Since they also have no hypotheses, every result that their proofs use is proved in Lean, in the two packages or in their dependencies. The upper bounds on the mean bias and on the variance in Theorems 4.10 and 4.11 and the variance–bias inequality enter the formal proofs as Lean theorems. The first and the last of these come from [7].

In their source packages, the four theorems also passed an independent check, which has three parts. First, we wrote out their statements a second time, with every definition that they involve, in separate files that import only Mathlib and none of our proofs. For the barrier, these are the definitions of the disorder law, the Hamiltonian, the Gibbs measure, the heat-bath kernels, the total variation distances, and the two quantities displayed above. For the fluctuations and the mean bias, they are the definitions on which `MainFluctuationStatement` rests, copied from OpenAI's sources at the commit cited above, and the definitions of the pressures $p_n(\beta)$ and of their limit. A comparison confirmed that each proved theorem has exactly the statement written out in this way. Second, the same comparison confirmed that the proofs use no axioms other than the three above. Third, the statements and proofs of the theorems were exported from the compiled packages, with every declaration on which they depend, including the proofs of all lemmas that they use. This export was checked again by three type checkers for Lean's logic: Lean's own kernel and the external checkers nanoda and con-ron. All three accepted the barrier statements and, in a separate comparison, both the fluctuation theorem and the mean-bias corollary. Declarations of the packages that these proofs do not use are not part of the check.

22.3 The formal and the written barrier proofs

The formal proof of Theorem 1.1 does not follow the proof in this paper in every step. The paper proves the local three-replica expansion of Proposition 7.13 with a remainder $O(h^4k^2)$, where h is the width parameter and k the relative imbalance of (7.1), by estimates

for smooth limiting flows. The formal proof instead works with finite chains of Gaussian integrations, using the hierarchical recursion of [77, Section 2] and its accompanying Lean development [83]. It proves the upper bound contained in (7.35) for these chains, at the ratio $\vartheta = 1/2$ used in the dynamical argument and on a smaller range of the relative imbalance k. Its remainder is of order $h^{7/2}k^2 + h^3|k|^3$ instead of h^4k^2, up to an additional error that tends to zero as the length of the chains tends to infinity. Only this upper bound enters the proof of the uniform cubic deficit in Proposition 7.15, where the remainder must be small compared with h^3k^2. The weaker remainder has this property when h and $|k|$ are small. Either estimate suffices for Theorem 1.1; they lead to different constants, not to a different exponent.

In the final counting step, the formal proof and Section 12 use the same kind of count. Only the first reference configuration of each block is counted, and the sign obstruction follows, as in Proposition 12.2, once a count of the form (12.2) is less than one. The formal proof passes through finite grids of threshold values, while Section 12 uses the uniform law of the thresholds directly. We conclude that the formal proof certifies Theorem 1.1 and Corollary 1.2, not every intermediate estimate in the paper's proof of them.

22.4 The formal proof of the two-sided bound

Our formal proof of Proposition 20.1 follows the route of [81] summarized in Section 20.2, and its mathematics is that of [81]. We do not claim a new proof of Proposition 20.1 or of Theorem 1.6, and we do not claim to have found or repaired a gap in the argument of [81]. What the formalization supplies is a complete treatment of the particular variational problems to which [81, Section 7] applies the general results of [81, Sections 3–6]. A paper proof can handle these problems by checking that they satisfy the hypotheses of general statements. In the formal proof, each such check is carried out in full detail. We now describe these problems and the cases that arise, in the notation of Section 20.2.

The comparisons and their minimizers. The upper bound in Proposition 20.1 rests on the comparisons of [81, Section 7.3]. For each $\varepsilon \in [\varepsilon_0, \beta^2)$ and $\theta \in [0,1]$, we minimize a functional of the form [81, (6)] over overlap quantiles r on the mass interval $[s,1]$. The lower radius is $r(s) = r_{\min} = \bar{u}_*/(2M(s))$, and the low-field perturbation of strength θ acts on $[0,s)$. The general results of [81] are stated for the comparison families of [81, Definition 6.1], which allow a positive lower radius $r_{\min}$ and an omitted lower interval. All overlap moments in them refer to minimizers of the perturbed paths. In the formal proof, we establish their hypotheses and conclusions for the minimizers of these particular functionals and for the finitely many successive comparisons built from them. For $\theta = 0$, and for the successive comparisons, whose unperturbed paths are nonnegative and nondecreasing, we do this throughout the strength interval $0 < \widehat{\lambda} \le d^2$. For $\theta > 0$, we use the two-sided linear bound only at the fixed final strength $\widehat{\lambda} = d^2$, uniformly

in ε, θ, and the minimizer. We use neither a pressure at the unperturbed paths nor an integration in the strength from zero. This restriction is necessary. Let $e_{\widehat{\lambda}}$ be the function e of Section 20.2 at strength $\widehat{\lambda}$. Since $e_{\widehat{\lambda}}(r_{\min}) = \widehat{\lambda}$ for large n, the perturbed field path is nondecreasing at s only if $\widehat{\lambda} r_{\min} \geq \theta c_L r_0$, that is, only if $\widehat{\lambda} \geq \theta c_L d^2/(2\varepsilon)$.

In the formal proof, we use the strength profile $e_{\widehat{\lambda}}(r) = \widehat{\lambda} \tilde{f}(\widehat{\lambda} r/d^3)$, where $\tilde{f}(t) = 1$ for $t \leq 1$ and $\tilde{f}(t) = (1 + \log t)/t$ for $t > 1$. This function has the monotonicity properties required of the profile f in [81, (6)]: its logarithmic slope $-t\tilde{f}'(t)/\tilde{f}(t)$ equals 0 for $t \leq 1$ and $\log t/(1 + \log t)$ for $t > 1$. It is continuously differentiable but not smooth at $t = 1$, and the parts of the argument that use the profile were carried out for this choice.

The three cases. Consider a minimizer at strength $\widehat{\lambda}$ and a mass $x \geq A_{\widehat{\lambda}}$ at which the two-sided linear bound fails, while it holds above $2x$, as in the propagation argument of Section 20.2. There are three cases.

1. *Comparable strength at a small scale*: $x \to 0$ and $e_{\widehat{\lambda}}(x)/x^2$ is bounded below. This case is excluded by the linear limit of [81, Theorem 5.1], which has to be proved for minimizers with the positive lower radius $r_{\min}$. Terms involving $r_{\min}$ enter as multiples of $\widehat{\lambda} r_{\min}$, and they are negligible at the cubic scale. To see this, recall from [81, (3) and (5)] that

$$\underline{x} = \frac{\sqrt{n}E}{d}, \qquad \bar{u}_* = \frac{1}{nE}, \qquad L = \min\Big\{\bar{u}_*, \frac{\underline{x}^3}{d^2}\Big\},$$

and that the upper bound on the derivative of the inverse quantile is

$$M(\bar{u}) = n^{\delta}\Big[\frac{\bar{u} + \bar{u}_*}{L} + \frac{d^3}{E}(\bar{u} + \bar{u}_*)^2\Big].$$

Since $M(s) \geq n^{\delta}\bar{u}_*/L$ and $r_{\min} = \bar{u}_*/(2M(s))$, and since $x \geq A_{\widehat{\lambda}} \geq n^{\delta}\underline{x} \geq \underline{x}$ for $\widehat{\lambda} \leq d^2$, we have

$$\widehat{\lambda} r_{\min} \leq \frac{\widehat{\lambda} L}{2n^{\delta}} \leq \frac{d^2 L}{2n^{\delta}} \leq \frac{\underline{x}^3}{2n^{\delta}} \leq \frac{x^3}{2n^{\delta}}.$$

We use the one-site cavity estimate of [81, Proposition 4.1] at the two scales x and $\bar{x} = \max\{x, \sqrt{e_{\widehat{\lambda}}(x)}\}$, and $\bar{x}/x$ may tend to infinity; for example, $\bar{x}/x = n^{5m-\delta}$ when $\widehat{\lambda} = d^2$ and $x = A_{d^2}$. The identification of the limiting scalar and overlap distributions [81, Lemma 5.4] and the first variations on gaps of the limiting support [81, Lemma 5.6] are carried out for these minimizers and their admissible competitors.

2. *Vanishing strength*: $e_{\widehat{\lambda}}(x)/x^2 \to 0$, either with $x \to 0$ or with x bounded below. The next comparison is built from the perturbed paths of this minimizer, as in [81, Lemma 6.2]. The alternative described in Section 20.2 is then followed for a fixed number $N = \lceil 10/\delta \rceil$ of successive comparisons [81, Lemmas 6.3–6.5 and Proposition 6.6]. At each stage, we exclude a violation with strength comparable to the scale as in the first or the third case.

3. *Macroscopic scale*: x and $e_{\widehat{\lambda}}(x)/x^2$ are bounded below, and $d^2 \le \beta^2 - c_*$ for a fixed $c_* > 0$. We exclude this case using the convergence of the minimizers to the quantile function of the Parisi measure [81, Proposition 5.9], applied to the comparison at hand.

The successive comparisons follow one of the two covariance alternatives of [81, Definition 6.1 and Corollary 6.7]. If $\varepsilon < \beta^2/4$, the perturbed matrix covariance is bounded below by a fixed positive constant, and they act on all masses. If $\varepsilon \ge \beta^2/4$, they act only on $[s, 1]$, with lower radius zero, and the matrix covariance has a jump of at least $\beta^2/4 - d^2$ at s. We then choose the coefficient c_0 so that the total matrix covariance Nc_0^2 removed by the successive comparisons is less than half of a fixed positive constant that bounds this jump from below; their perturbed matrix paths therefore remain nondecreasing at s. For $\theta > 0$, we establish the two-sided linear bound at the fixed final strength d^2 [81, Lemma 7.2] and the bound $S_{\bar{u}} \le Cd^3/\varepsilon$ on $[0, s)$ [81, Lemma 7.3], uniformly in ε and θ. These two estimates enter the integration in θ of [81, Lemma 7.4]. As in [81], the constants in the two-sided linear bound, the depth N, and the coefficient c_0 are fixed before $n \to \infty$. The final threshold in n is uniform in ε, θ, and the minimizer, and in $\widehat{\lambda}$ over the strengths at which each comparison is used.

22.5 The upper bound

The proposition `SKConstants.OptimizedSKTypicalStartStatement`, defined in `SKConstants/SKModel.lean`, states that, for every $\beta > 1$ and every $\omega \in (0, 1/4)$,

$$\mathbb{P}_{J, X_0 \sim \pi^J}\left\{ d^J\left(X_0, \exp(n^{1-\omega})\right) > 1/4 \right\} \longrightarrow 0.$$

The declaration uses c for the parameter denoted ω here, with both β and c fixed before $n \to \infty$.

In the formal model, the variables g_{ij} are independent standard Gaussians, and the completed matrix and Hamiltonian are

$$W_{ij} = \frac{g_{ij} + g_{ji}}{\sqrt{2n}}, \qquad H_g^c(x) = \frac{1}{2} x^{\mathsf{T}} W x,$$

including its diagonal, with β outside the Hamiltonian. For $i < j$, we define $J_{ij} = \sqrt{n}\, W_{ij}$. These couplings are independent standard Gaussians, and

$$H_g^c(x) = H_n^J(x) + \frac{1}{2} \operatorname{tr} W, \qquad Z_n^c = e^{\beta \operatorname{tr} W/2} Z_n.$$

The trace term is independent of x and cancels in the Gibbs normalization, as in Lemma 3.1(b), and in every energy difference. The Gibbs probabilities and flip rates coincide with those of Section 3. The same holds for the semigroup and the event defined

by the total variation distance from a fixed initial spin. The formal generator is $n(K^J - I)$, so each site has rate one, as in (3.12). Its joint law samples the completed matrix and then a Gibbs spin. By integrating out the independent diagonal, we obtain the joint law of Definition 3.3. The primitive representation in Section 13.1 uses the same completed matrix law. The initial spin is fixed before its transition law and total variation distance are formed. By the complementary form of Lemma 3.4, the displayed limit is Theorem 1.3. The formal sequence is also defined at sizes zero and one, which does not affect the limit.

The theorem `SKConstants.Completion.Main.optimizedSKTypicalStartStatement_holds` proves this proposition without additional hypotheses. The original upper-bound package at commit `352426b9007b6bbb31a3c8cfe6887caa0badf3fd` was checked on 19 October 2026, using Lean version `v4.35.0-rc2` and Mathlib at commit `065356127b1dc0016f66b7283ce0ce2c4055aa55`. A fresh build succeeded at this revision. The package's `Challenge.lean` contains 47 independently stated claims and imports only Mathlib. Its definitions explicitly include the completed disorder law, Gibbs kernel, semigroup, total variation distance, and joint probability. The official Comparator confirmed equality of the proved statements and these independent statements, including their definitions. The axiom audit of all 20,281 exported development declarations in 1,605 modules found only `propext`, `Classical.choice`, and `Quot.sound`; the main theorem has exactly these three axiom dependencies. The same Comparator run checked all 47 claims and their proof dependencies with Lean's kernel, NanoDa, and con-ron. All three accepted them. The upper-bound theorem is also one of the five final statements checked in the combined package described at the start of Section 22.

The formal proof follows the route of Part IV. It constructs Brownian motion, proves the scalar and Gaussian-query laws, establishes the fine entropy and sharp coarse estimates, and combines coverage and the connecting sequence with the finite-chain mixing reduction. The steps are organized as separately proved statements and assembly theorems, whose hypotheses are supplied by the final proof. The parameter choice used by that proof is described in Section 22.6. The development includes Ezzeri Esa's Stochastic-Calculus library and adapted OpenAI Lean modules, both under the Apache-2.0 license. Intermediate estimates of the written proof are checked only insofar as they enter the formal proof or are separately stated formal results.

22.6 Scope and availability

The combined check concerns the five final statements described above; the source-package checks of Section 22.2 concern the four lower-bound and fluctuation results described there. Intermediate results in their proofs, such as Proposition 20.1, are checked only as parts of these proofs. The quantitative and almost-sure forms in Theorem 12.6 are not formally verified. These are the bounds on the probability of the exceptional

disorders and on the Gibbs mass of the exceptional initial states, the lower bounds on the distance that tend to $1/2$, the extension to the Metropolis dynamics and to other single-site dynamics with bounded rates, the bound on the relaxation time, and the almost-sure statements for arbitrary couplings of the disorders across system sizes. The sharper local expansion of Proposition 7.13 is not formally verified either; see Section 22.3. The proposition `MainFluctuationStatement` concerns exponents only, and the limit law and the positive variance prefactor claimed in [82] are not part of it. The proposition `LowerBoundStatement` contains the loss $n^{-\varepsilon}$ for an arbitrary $\varepsilon > 0$, and the bound (21.2), which would follow from a positive variance prefactor, is not part of it.

Corollaries 1.4 and 1.5 are not separately formalized. The development proves admissibility of the affine family in Proposition 19.2 throughout $0 < \varepsilon_\omega < 1/80$, including the choice $\varepsilon_\omega = (1/4 - \omega)/20$. The final proof of Theorem 1.3 instead uses the same affine family at $\min\{1/2000, (1/4 - \omega)/20\}$ and proves the target for every $\omega \in (0, 1/4)$. A separate formal result characterizes the open range of the development's strict estimate system, and another checks an explicit rational instance. The ancillary calculation described after Proposition 19.2, including every identity for Table 19.1, is not a separately certified formal statement.

OpenAI's Lean sources [83] and the development [7] are public. The combined sources, independent `Challenge.lean`, and verification records are available in `https://github.com/pmaronow/fluctuations-palomar`; the records identify the checked proof-source revision and the scope of each check. The original two source packages and the historical upper-bound package are maintained in the private repository `https://github.com/pmaronow/fluctuations-temp`. The latter was checked at commit `352426b9007b6bbb31a3c8cfe6887caa0badf3fd`, in its `glauber/lean` directory. The historical records remain available from the authors.

SECTION 23

Open problems

We collect some questions that this paper leaves open. The first four concern the dynamics, and the last two concern the law of F_n and the mean bias. We use the mixing times from a fixed configuration with threshold $1/4$ defined in (3.14).

The exponent of the typical mixing time. By Corollary 1.5, for every fixed $\beta > 1$ and every fixed $\omega \in (0, 1/4)$, the realized initial configuration $X_0 \sim \pi^J$ satisfies $e^{A_\beta n^{7/30}} < T^J(X_0) \le e^{n^{1-\omega}}$ with probability tending to one. The same bounds hold in discrete time, with the integer parts of the two time horizons. We do not know whether there exists an exponent ζ such that $\log T^J(X_0) = n^{\zeta + o_{\mathbb{P}}(1)}$, in the sense of Section 3.5. If it exists, then $7/30 \le \zeta \le 3/4$. The numerical and analytical work cited in Section 1.3 suggests that $\zeta = 1/3$ [18, 19, 65, 76]. Billoire also studied the distribution of the logarithm of the relaxation time over samples of the disorder [18]. In the same way, one may ask for the fluctuations of $\log T^J(X_0)$ over J and X_0.

The limits of the present argument. In the proof of Theorem 1.1, the upper bound on the mean bias is used to make the cubic deficit, of order ρ^3 at the scale $\rho = Rn^{-\lambda}$, dominate $\Delta_{n,\beta}$. Suppose more generally that $\Delta_{n,\beta} \le C_\beta n^{-\gamma}$ for some fixed $\gamma > 0$, and retain the variance bound $\mathrm{Var}(F_n) \le C_v n^{7/15}$. We set the tolerance of the overlap constraints at fixed scales to $\eta_n = n^{-\xi}$, where the exponents (λ, ξ) satisfy (10.10) and $3\lambda \le \gamma$; that is,

$$0 < \lambda \le \gamma/3, \qquad \lambda < 1/4, \qquad \lambda < \xi < \min\{1/4, 3\lambda/2\}.$$

We check that the conditions on the scales in the proof of Theorem 1.1 still hold. Indeed, the interval for ξ is nonempty; it ensures that $\eta_n = o(\rho)$, $\rho^3 = o(\eta_n^2)$, and $n\rho^4/\log n$, $n\eta_n^4/\log n \to \infty$; and the Chebyshev bound for the denominator vanishes since $\lambda < 1/4 < 23/90$. Choosing R with a fixed margin when $3\lambda = \gamma$, we obtain the time $\exp(An^\lambda)$ for some $A > 0$. The barrier exponent is therefore one third of the exponent of the upper bound on the mean bias, as long as this third is below $1/4$: the case $0 < \gamma < 3/4$ permits $\lambda = \gamma/3$, while $\gamma \ge 3/4$ permits every fixed $\lambda < 1/4$. In the latter case, choosing a slightly larger admissible power and using the contraction of total variation also gives the time $\exp(n^\lambda)$ with coefficient one. The upper bound $n^{-7/10}$ of [6, Theorem 1.1] is the case $\gamma = 7/10$, and we obtain the exponent $\lambda = 7/30$ of Theorem 1.1. The exponent $1/4$ itself is not reached: at $\rho = Rn^{-1/4}$, the quantity $n\rho^4$ is bounded, and the union bound over polynomially many constrained partition functions in the disorder

event fails. The prediction $\Delta_{n,\beta} \approx n^{-3/4}$ of Franz, Parisi, and Ricci-Tersenghi [56] lies exactly at this threshold. We expect that the exponent $1/4$, and the value $1/3$ suggested by physics, can be reached only with new ideas, such as overlap constraints with a larger pressure deficit or bounds on constrained partition functions that hold simultaneously without a union over constraints.

For the upper bound, the sufficient conditions of Proposition 19.2 require

$$\omega < \mathsf{q} < \mathsf{s}_{\mathrm{b}} < \mathsf{s}_{\mathrm{seq}} < \mathsf{e} < \mathsf{u}_{\mathrm{E}} < \mathsf{k}_{\mathrm{f}}/2 \le 1/4.$$

Both arguments stop at the exponent $1/4$. We do not know whether this coincidence has a common explanation.

Dependence on the temperature. The coefficient A_β depends on β through the scalar Parisi problem, through the constant in the upper bound on the mean bias of [6], and through the constant in the upper bound on the variance of [8]. The estimates used here do not control these constants uniformly as $\beta \downarrow 1$ or $\beta \to \infty$. At $\beta = 1$, the mixing time is $n^{2/3+o(1)}$ [78, Theorem 1.1]. Billoire's simulations with binary couplings suggest an approximate collapse of the centered variable $\log \tau_J - \overline{\log \tau}$ after division by $n^{1/3}(\beta - 1)$, with visible corrections near the transition [18, equation (6), Figure 3 and the following discussion]. Here τ_J is an equilibrium autocorrelation time that depends on the disorder sample, and $\overline{\log \tau}$ is the disorder average of $\log \tau_J$. We ask for a description of the crossover between the two regimes in a window of temperatures around $\beta = 1$, and for bounds that remain effective as $\beta \to \infty$.

Worst-case and typical initial configurations. Sellke proved that the mixing time from the worst initial configuration is exponential in n when β exceeds an absolute constant [99, Theorem 1.1]. By Theorem 1.3, the mixing time from a typical equilibrium configuration is at most $\exp(n^{1-\omega})$ for every fixed $\omega \in (0, 1/4)$, with joint probability tending to one, whereas Sellke's worst-start bound is exponential in n at sufficiently large β. For every $\beta > 1$, Theorem 12.6 implies that, with probability tending to one, the relaxation time of the heat-bath and Metropolis dynamics is at least $e^{An^{7/30}}/(Cn^{7/30})$, so that their spectral gaps are at most $Cn^{7/30}e^{-An^{7/30}}$. Is the spectral gap exponentially small in n for every $\beta > 1$, and is the mixing time from the worst initial configuration exponential in n? The sufficient hypotheses of [16, Theorem 1.9] include the vanishing of the second derivative at zero of the covariance function, and exclude the quadratic covariance of SK. Their general overlap-barrier criterion [16, Theorem 1.3] remains available if its barrier hypothesis can be verified. A second question concerns the size of the distance. By Theorem 12.6, the distance from equilibrium of the chain started from most initial configurations is at least $1/2 - Ce^{-cn^{7/30}}$ at every time $t \le e^{An^{7/30}}$. The value $1/2$ comes from spin-flip symmetry, which gives Gibbs mass at least $1/2$ to every half-space $\{y : R(v, y) \le 0\}$, and it is the limit of the argument. Does the conclusion of Theorem 1.1 hold for every fixed $\tau < 1$?

The limiting law of F_n. Theorem 1.6 identifies the exponent of the fluctuations but not their law. OpenAI states that $n^{-1/3}\operatorname{Var}(F_n)$ converges to a positive constant $a_2(\beta)$, and that $(F_n - \mathbb{E}F_n)/\sqrt{\operatorname{Var}(F_n)}$ converges in law to a nondegenerate limit ν_β [82, Theorems 1.1 and 1.2]. We do not use these statements. The limit ν_β is not identified with a known distribution. Parisi and Rizzo predicted upward deviations of F_n/n with speed n and power $6/5$, and downward deviations with speed n^2 and power $12/5$ [94, equations (12)–(15)]. If these tails match the typical fluctuations, the right and left tails of ν_β should be of the forms $\exp(-cx^{6/5})$ and $\exp(-cx^{12/5})$. This conditional passage from large-deviation tails to a typical scale follows the matching argument of [93, Section VI]. On the rigorous side, [8, Theorem 1.5] gives two-sided bounds with the power $6/5$ on the probability that F_n/n exceeds $P_\beta + x$, for x in a range $C_\beta n^{-2/3} \le x \le x_\beta$. The lower end of this range is of the order of the upper bound on the mean bias in [8, Theorem 1.4]. Bounds of this type for deviations from $\mathbb{E}F_n/n$, down to the fluctuation scale $n^{-5/6}$, would make the matching argument of Parisi and Rizzo rigorous for upward deviations.

The exponent of the mean bias. By Corollary 1.7, we have $\Delta_{n,\beta} \ge c_{\beta,\varepsilon} n^{-5/6-\varepsilon}$ for every fixed $\varepsilon > 0$, and by [6, Theorem 1.1], we have $\Delta_{n,\beta} \le C_b n^{-7/10}$. If $\Delta_{n,\beta} = n^{-a+o(1)}$ for some a, then $7/10 \le a \le 5/6$. Franz, Parisi, and Ricci-Tersenghi predict $a = 3/4$, possibly with logarithmic corrections [56]. The lower bound cannot be improved by the variance–bias inequality (1.7) alone. Since $\operatorname{Var}(F_n) = n^{1/3+o(1)}$, this inequality gives no lower bound with an exponent smaller than $5/6$, and such a bound would require a different argument. On the upper side, the discussion in [6, Section 1.4] explains which features of the proof of [6, Theorem 1.1] account for the difference between $7/10$ and $3/4$. The estimates there do not control the constants uniformly as $\beta \to \infty$, so the corresponding question for the ground-state energy is also open.

References

[1] Michael Aizenman, Joel L. Lebowitz, and David Ruelle. Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Communications in Mathematical Physics*, 112(1):3–20, 1987. doi: 10.1007/BF01217677. https://doi.org/10.1007/BF01217677.

[2] Michael Aizenman, Robert Sims, and Shannon L. Starr. Extended variational principle for the Sherrington–Kirkpatrick spin-glass model. *Physical Review B*, 68: 214403, 2003. doi: 10.1103/PhysRevB.68.214403.

[3] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1418–1430. ACM, 2022. doi: 10.1145/3519935.3520048. Theorem locators refer to the preprint arXiv:2106.04105v2.

[4] Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-down in localization schemes and applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1094–1105. ACM, 2024. doi: 10.1145/3618260.3649622. URL https://arxiv.org/abs/2407.16104v1. Statement locators use https://arxiv.org/abs/2407.16104v1.

[5] P. M. Aronow and Patrick Lopatto. Overlap structure of mixed even p-spin models. https://arxiv.org/abs/2609.17442v1, 2026. URL https://arxiv.org/abs/2609.17442v1. Version 1, September 15, 2026.

[6] P. M. Aronow and Patrick Lopatto. An $N^{-7/10}$ bound for the finite-size bias of the Sherrington–Kirkpatrick free energy. Hexagon preprint, https://hexagonmath.org/pdf/2610.00177v1, October 2026. URL https://hexagonmath.org/pdf/2610.00177v1. Version 1; publication date October 12, 2026. hexagon:2610.00177v1.

[7] P. M. Aronow and Patrick Lopatto. Finite-size mean bias in the Sherrington–Kirkpatrick model. Lean development, https://github.com/pmaronow/sk-mean-bias-palomar/tree/7601a9d5ade08066632bc68b9a615e163f2dd02c, 2026. URL https://github.com/pmaronow/sk-mean-bias-palomar/tree/7601a9d5ade08066632bc68b9a615e163f2dd02c. Commit 7601a9d5ade08066632bc68b9a615e163f2dd02c.

[8] P. M. Aronow and Patrick Lopatto. Quantitative Parisi formulas and fluctuations in the Sherrington–Kirkpatrick model. https://arxiv.org/abs/2609.09103v1, 2026. URL https://arxiv.org/abs/2609.09103v1. Version 1, September 8, 2026.

[9] T. Aspelmeier and M. A. Moore. Free-energy barriers in the Sherrington–Kirkpatrick model. *Physical Review E*, 105:034138, 2022. doi: 10.1103/PhysRevE.105.034138.

[10] Timo Aspelmeier. An exact relation between free energy fluctuations and bond chaos in the Sherrington–Kirkpatrick model. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(4):P04018, 2008. doi: 10.1088/1742-5468/2008/04/P04018. Equation locators refer to arXiv:0802.3007v1.

[11] Timo Aspelmeier, Alain Billoire, Enzo Marinari, and Michael A. Moore. Finite size corrections in the Sherrington–Kirkpatrick model. *Journal of Physics A: Mathematical and Theoretical*, 41(32):324008, 2008. doi: 10.1088/1751-8113/41/32/324008. Section locators refer to arXiv:0711.3445v1.

[12] Antonio Auffinger and Wei-Kuo Chen. On properties of Parisi measures. *Probability Theory and Related Fields*, 161:817–850, 2015. doi: 10.1007/s00440-014-0563-y. URL `https://arxiv.org/abs/1303.3573v1`. Statement locators use `https://arxiv.org/abs/1303.3573v1`.

[13] Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Communications in Mathematical Physics*, 335:1429–1444, 2015. doi: 10.1007/s00220-014-2254-z. URL `https://arxiv.org/abs/1402.5132v3`. Statement locators use `https://arxiv.org/abs/1402.5132v3`.

[14] Jinho Baik and Ji Oon Lee. Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model. *Journal of Statistical Physics*, 165(2):185–224, 2016. doi: 10.1007/s10955-016-1610-0. Statement locators refer to arXiv:1505.07349v5.

[15] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019. doi: 10.1016/j.jfa.2019.01.007. Statement locators refer to arXiv:1712.03676v2.

[16] Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Communications in Mathematical Physics*, 361(1):1–52, 2018. doi: 10.1007/s00220-018-3152-6. URL `https://arxiv.org/abs/1705.04243v3`. Statement locators use `https://arxiv.org/abs/1705.04243v3`, version 3, March 2, 2018.

[17] Alain Billoire. Numerical estimate of the finite-size corrections to the free energy of the Sherrington–Kirkpatrick model using Guerra–Toninelli interpolation. *Physical Review B*, 73:132201, 2006. doi: 10.1103/PhysRevB.73.132201.

[18] Alain Billoire. Distribution of timescales in the Sherrington–Kirkpatrick model. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(11):P11034, 2010. doi: 10.1088/1742-5468/2010/11/P11034. Equation and figure locators refer to arXiv:1003.6086v3.

[19] Alain Billoire and Enzo Marinari. Correlation time scales in the Sherrington–Kirkpatrick model. *Journal of Physics A: Mathematical and General*, 34:L727–L734, 2001. doi: 10.1088/0305-4470/34/50/101.

[20] Elmar Bittner and Wolfhard Janke. Free-energy barriers in the Sherrington–Kirkpatrick model. *Europhysics Letters*, 74(2):195–201, 2006. doi: 10.1209/epl/i2006-10007-y.

[21] Mathews Boban, Anqi Li, and Shayan Oveis Gharan. Rank-1-perturbed trickledown theorems: mixing time of Glauber dynamics for the Sherrington–Kirkpatrick model up to $\beta \leq \frac{1}{2} + \varepsilon$. `https://arxiv.org/abs/2609.13138v1`, 2026. URL `https://arxiv.org/abs/2609.13138v1`. Version 1.

[22] Stefan Boettcher. Extremal optimization for Sherrington–Kirkpatrick spin glasses. *The European Physical Journal B*, 46(4):501–505, 2005. doi: 10.1140/epjb/e2005-00280-6.

[23] Stefan Boettcher. Simulations of ground state fluctuations in mean-field Ising spin glasses. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(7):P07002, 2010. doi: 10.1088/1742-5468/2010/07/P07002. Section locators refer to arXiv:0906.1292v2.

[24] V. I. Bogachev, A. V. Kolesnikov, and K. V. Medvedev. Triangular transformations of measures. *Sbornik: Mathematics*, 196(3):309–335, 2005. doi: 10.1070/SM2005v196n03ABEH000882.

[25] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325:333–366, 2014. doi: 10.1007/s00220-013-1862-3.

[26] Jean-Philippe Bouchaud, Leticia F. Cugliandolo, Jorge Kurchan, and Marc Mézard. Out of equilibrium dynamics in spin-glasses and other glassy systems. `https://arxiv.org/abs/cond-mat/9702070v1`, 1997. URL `https://arxiv.org/abs/cond-mat/9702070v1`. Version 1.

[27] Jean-Philippe Bouchaud, Florent Krzakala, and Olivier C. Martin. Energy exponents and corrections to scaling in Ising spin glasses. *Physical Review B*, 68:224404, 2003. doi: 10.1103/PhysRevB.68.224404.

[28] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL `https://academic.oup.com/book/26549`.

[29] Anton Bovier and Anton Klimovsky. The Aizenman–Sims–Starr and Guerra's schemes for the SK model with multidimensional spins. *Electronic Journal of Probability*, 14(8):161–241, 2009. doi: 10.1214/EJP.v14-611.

[30] Herm Jan Brascamp and Elliott H. Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976. doi: 10.1016/0022-1236(76)90004-5. URL `https://www.scienced`

irect.com/science/article/pii/0022123676900045. https://doi.org/10.1016/0022-1236(76)90004-5.

[31] A. J. Bray and M. A. Moore. Metastable states in the solvable spin glass model. *Journal of Physics A: Mathematical and General*, 14(9):L377–L383, 1981. doi: 10.1088/0305-4470/14/9/012.

[32] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. doi: 10.1002/cpa.3160440402. URL https://www.ceremade.dauphine.fr/~carlier/Brenier91.pdf.

[33] Sourav Chatterjee. Disorder chaos and multiple valleys in spin glasses. https://arxiv.org/abs/0907.3381v4, 2009. URL https://arxiv.org/abs/0907.3381v4. Version 4, 23 November 2009.

[34] Sourav Chatterjee. A general method for lower bounds on fluctuations of random variables. *Annals of Probability*, 47(4):2140–2171, 2019. doi: 10.1214/18-AOP1304. URL https://arxiv.org/abs/1706.04290v3. Statement locators use https://arxiv.org/abs/1706.04290v3.

[35] Hong-Bin Chen, Alice Guionnet, Justin Ko, Bertrand Lacroix-A-Chez-Toine, and Jean-Christophe Mourrat. One-sided large deviations for the ground-state energy of spin glasses. https://arxiv.org/abs/2603.06368v1, 2026. URL https://arxiv.org/abs/2603.06368v1. Version 1, March 6, 2026.

[36] Wei-Kuo Chen. Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *The Annals of Probability*, 45(6A):3929–3966, 2017. doi: 10.1214/16-AOP1154. URL https://arxiv.org/abs/1501.06635v2. Statement locators use https://arxiv.org/abs/1501.06635v2.

[37] Wei-Kuo Chen. A Gaussian convexity for logarithmic moment generating functions with applications in spin glasses. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 62(1):195–206, 2026. doi: 10.1214/24-AIHP1527. URL https://doi.org/10.1214/24-AIHP1527. Preprint version arXiv:2311.08351v2.

[38] Wei-Kuo Chen and Wai-Kit Lam. Order of fluctuations of the free energy in the SK model at critical temperature. *ALEA. Latin American Journal of Probability and Mathematical Statistics*, 16(1):809–816, 2019. doi: 10.30757/ALEA.v16-29. URL https://alea.impa.br/articles/v16/16-29.pdf. https://doi.org/10.30757/ALEA.v16-29.

[39] Wei-Kuo Chen, Partha Dey, and Dmitry Panchenko. Fluctuations of the free energy in the mixed p-spin models with external field. *Probability Theory and Related Fields*, 168(1–2):41–53, 2017. doi: 10.1007/s00440-016-0705-5. URL https://arxiv.org/abs/1509.07071v2. Statement locators use https://arxiv.org/abs/1509.07071v2.

[40] S. G. W. Colborne. A Monte Carlo study of the dynamics of the Ising SK model. *Journal of Physics A: Mathematical and General*, 23(17):4013–4023, 1990. doi: 10.1088/0305-4470/23/17/030.

[41] Elizabeth W. Collins-Woodfin and Han Gia Le. Free energy fluctuations in SK and related spin glass models: A literature survey. `https://arxiv.org/abs/2510.26960v3`, 2025. URL `https://arxiv.org/abs/2510.26960v3`. Version 3, November 7, 2025.

[42] Francis Comets and Jacques Neveu. The Sherrington–Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Communications in Mathematical Physics*, 166(3):549–564, 1995. doi: 10.1007/BF02099887. URL `https://doi.org/10.1007/BF02099887`. `https://doi.org/10.1007/BF02099887`.

[43] Pierluigi Contucci and Emanuele Mingione. A multi-scale spin-glass mean-field model. *Communications in Mathematical Physics*, 368(3):1323–1344, 2019. doi: 10.1007/s00220-019-03308-8.

[44] Andrea Crisanti and Tommaso Rizzo. Analysis of the ∞-replica symmetry breaking solution of the Sherrington–Kirkpatrick model. *Physical Review E*, 65:046137, 2002. doi: 10.1103/PhysRevE.65.046137. Equation locators refer to arXiv:cond-mat/0111037v5.

[45] Andrea Crisanti, Giovanni Paladin, Hans-Jürgen Sommers, and Angelo Vulpiani. Replica trick and fluctuations in disordered systems. *Journal de Physique I*, 2(7):1325–1332, 1992. doi: 10.1051/jp1:1992213.

[46] Leticia F. Cugliandolo and Jorge Kurchan. On the out-of-equilibrium relaxation of the Sherrington–Kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27:5749–5772, 1994. doi: 10.1088/0305-4470/27/17/011.

[47] Cirano De Dominicis and Imre Kondor. Eigenvalues of the stability matrix for Parisi solution of the long-range spin-glass. *Physical Review B*, 27(1):606–608, 1983. doi: 10.1103/PhysRevB.27.606.

[48] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37. URL `https://doi.org/10.1007/978-3-030-79876-5_37`. `https://doi.org/10.1007/978-3-030-79876-5_37`.

[49] Partha S. Dey and Taegu Kang. Fluctuations for the Sherrington–Kirkpatrick spin glass model near the critical temperature. `https://arxiv.org/abs/2603.05636v1`, 2026. URL `https://arxiv.org/abs/2603.05636v1`. Version 1, March 5, 2026.

[50] Partha S. Dey and Qiang Wu. Mean field spin glass models under weak external field. *Communications in Mathematical Physics*, 402(2):1205–1258, 2023. doi: 10.100

7/s00220-023-04742-5. URL https://arxiv.org/abs/2112.11349v3. Statement locators use https://arxiv.org/abs/2112.11349v3.

[51] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, 6(3):695–750, 1996. doi: 10.1214/aoap/1034968224.

[52] Hang Du and Brice Huang. Fluctuations of the Sherrington–Kirkpatrick free energy at critical temperature. https://arxiv.org/abs/2607.02172v2, 2026. URL https://arxiv.org/abs/2607.02172v2. Version 2, 8 July 2026.

[53] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington–Kirkpatrick gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022. doi: 10.1109/FOCS54457.2022.00038. URL https://arxiv.org/abs/2203.05093v2. Numbered statements here use the extended preprint https://arxiv.org/abs/2203.05093v2, version 2, February 15, 2024.

[54] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature Ising models. *Probability Theory and Related Fields*, 182(3–4):1035–1051, 2022. doi: 10.1007/s00440-021-01085-x. URL https://arxiv.org/abs/2007.08200v2. Statement locators use https://arxiv.org/abs/2007.08200v2.

[55] Silvio Franz and Giorgio Parisi. Recipes for metastable states in spin glasses. *Journal de Physique I*, 5(11):1401–1415, 1995. doi: 10.1051/jp1:1995201.

[56] Silvio Franz, Giorgio Parisi, and Federico Ricci-Tersenghi. Overlap locking and nonperturbative effects in spin glasses. *Proceedings of the National Academy of Sciences*, 123(25):e2606171123, 2026. doi: 10.1073/pnas.2606171123. URL https://arxiv.org/abs/2602.18399v2. Statement and equation locators use https://arxiv.org/abs/2602.18399v2.

[57] Reza Gheissari and Aukosh Jagannath. On the spectral gap of spherical spin glass dynamics. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):756–776, 2019. doi: 10.1214/18-AIHP897. Statement locators refer to arXiv:1608.06609v3.

[58] Roy J. Glauber. Time-dependent statistics of the Ising model. *Journal of Mathematical Physics*, 4(2):294–307, 1963. doi: 10.1063/1.1703954. URL https://doi.org/10.1063/1.1703954. https://doi.org/10.1063/1.1703954.

[59] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics*, 233:1–12, 2003. doi: 10.1007/s00220-002-0773-5. URL https://link.springer.com/article/10.1007/s00220-002-0773-5. https://doi.org/10.1007/s00220-002-0773-5.

[60] Francesco Guerra and Fabio L. Toninelli. The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics*, 230(1):71–79, 2002. doi: 10.1007/s00220-002-0699-y. URL https://arxiv.org/abs/cond-mat/0204280v1. Statement locators use https://arxiv.org/abs/cond-mat/0204280v1.

[61] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1–2): 134–139, 1918. doi: 10.1093/biomet/12.1-2.134. URL https://academic.oup.com/biomet/article-abstract/12/1-2/134/193428.

[62] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, 2016. doi: 10.1090/proc/12968. URL https://arxiv.org/abs/1502.04398v4. Statement locators use https://arxiv.org/abs/1502.04398v4.

[63] Aukosh Jagannath and Ian Tobasco. Some properties of the phase diagram for mixed p-spin glasses. *Probability Theory and Related Fields*, 167:615–672, 2017. doi: 10.1007/s00440-015-0691-z. URL https://arxiv.org/abs/1504.02731v2. Statement locators use https://arxiv.org/abs/1504.02731v2.

[64] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 1991. doi: 10.1007/978-1-4612-0949-2.

[65] Harald Kinzelbach and Heinz Horner. Dynamics of the finite SK spin-glass. *Zeitschrift für Physik B Condensed Matter*, 84:95–114, 1991. doi: 10.1007/BF01453763.

[66] Imre Kondor. Parisi's mean-field solution for spin glasses as an analytic continuation in the replica number. *Journal of Physics A: Mathematical and General*, 16(4): L127–L131, 1983. doi: 10.1088/0305-4470/16/4/006. https://doi.org/10.1088/0305-4470/16/4/006.

[67] Gregory F. Lawler and Alan D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Transactions of the American Mathematical Society*, 309(2):557–580, 1988. doi: 10.1090/S0002-9947-1988-0930082-9.

[68] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, second edition, 2017. doi: 10.1090/mbk/107. URL https://pages.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf. With contributions by Elizabeth L. Wilmer.

[69] Patrick Lopatto. Full replica symmetry breaking in the Sherrington–Kirkpatrick model. https://arxiv.org/abs/2607.11756v3, 2026. URL https://arxiv.org/abs/2607.11756v3. Version 3, July 16, 2026.

[70] Gábor Lugosi and Gergely Neu. Generalization bounds via convex analysis. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on*

Learning Theory, volume 178 of *Proceedings of Machine Learning Research*, pages 3524–3546. PMLR, 2022. URL https://proceedings.mlr.press/v178/lugosi22a.html.

[71] N. D. Mackenzie and A. P. Young. Lack of ergodicity in the infinite-range Ising spin-glass. *Physical Review Letters*, 49:301–304, 1982. doi: 10.1103/PhysRevLett.49.301.

[72] N. D. Mackenzie and A. P. Young. Statics and dynamics of the infinite-range Ising spin glass model. *Journal of Physics C: Solid State Physics*, 16(27):5321–5337, 1983. doi: 10.1088/0022-3719/16/27/015.

[73] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323, 1995. doi: 10.1215/S0012-7094-95-08013-2. URL https://www.math.toronto.edu/~mccann/papers/monotone.pdf.

[74] Marc Mézard, Giorgio Parisi, and Miguel A. Virasoro. SK model: The replica solution without replicas. *Europhysics Letters*, 1(2):77–82, 1986. doi: 10.1209/0295-5075/1/2/006.

[75] Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, 54(4):FOCS19–1–FOCS19–38, 2025. doi: 10.1137/20M132016X. First published online 2021. Numbered section locators refer to https://arxiv.org/abs/1812.10897v2.

[76] Cécile Monthus and Thomas Garel. An eigenvalue method for computing the largest relaxation time of disordered systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(12):P12017, 2009. doi: 10.1088/1742-5468/2009/12/P12017.

[77] OpenAI. Stretched-exponential barriers for typical SK initial states. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Stretched-exponential-barriers-for-typical-SK-initial-states-September-24-2026/paper.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Stretched-exponential-barriers-for-typical-SK-initial-states-September-24-2026/paper.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[78] OpenAI. Critical mixing in the Sherrington–Kirkpatrick model. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Critical-mixing-in-the-Sherrington-Kirkpatrick-model-September-25-2026/paper.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Critical-mixing-in-the-Sherrington-Kirkpatrick-model-September-25-2026/paper.pdf. September 25, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[79] OpenAI. Critical slowing down in the Sherrington–Kirkpatrick model. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc7

29c18d98444fed42fd0529eeeb/preprints/Critical-slowing-down-in-the-Sherrington-Kirkpatrick-model-September-24-2026/paper.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Critical-slowing-down-in-the-Sherrington-Kirkpatrick-model-September-24-2026/paper.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[80] OpenAI. Cutoff throughout the high-temperature Sherrington–Kirkpatrick phase. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Cutoff-throughout-the-high-temperature-Sherrington-Kirkpatrick-phase-September-24-2026/paper.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/Cutoff-throughout-the-high-temperature-Sherrington-Kirkpatrick-phase-September-24-2026/paper.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[81] OpenAI. The low-temperature Sherrington–Kirkpatrick fluctuation scale. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/The-low-temperature-Sherrington-Kirkpatrick-fluctuation-scale-September-24-2026/The-low-temperature-Sherrington-Kirkpatrick-fluctuation-scale-September-24-2026.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/The-low-temperature-Sherrington-Kirkpatrick-fluctuation-scale-September-24-2026/The-low-temperature-Sherrington-Kirkpatrick-fluctuation-scale-September-24-2026.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[82] OpenAI. The low-temperature Sherrington–Kirkpatrick free-energy limiting law. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/The-low-temperature-Sherrington-Kirkpatrick-free-energy-limiting-law-September-24-2026/The-low-temperature-Sherrington-Kirkpatrick-free-energy-limiting-law-September-24-2026.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/The-low-temperature-Sherrington-Kirkpatrick-free-energy-limiting-law-September-24-2026/The-low-temperature-Sherrington-Kirkpatrick-free-energy-limiting-law-September-24-2026.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[83] OpenAI. Lean formalizations in the OpenAI mathematics repository. https://github.com/openai/math/tree/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/lean, 2026. URL https://github.com/openai/math/tree/fd4aeeb2ee4fc729c18d98444fed42fd0529ee

eb/lean. Commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb; directories lean/OAI/Probability/SKBarriers and lean/OAI/Probability/SKFluctuations, and lean/docs/227.md.

[84] OpenAI. A spectral gap throughout the high-temperature Sherrington–Kirkpatrick phase. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/A-spectral-gap-throughout-the-high-temperature-Sherrington-Kirkpatrick-phase-September-24-2026/main.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/A-spectral-gap-throughout-the-high-temperature-Sherrington-Kirkpatrick-phase-September-24-2026/main.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[85] OpenAI. A typical-start upper bound for low-temperature SK Glauber dynamics. OpenAI Math Release preprint, https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/A-typical-start-upper-bound-for-low-temperature-SK-Glauber-dynamics-September-24-2026/paper.pdf, September 2026. URL https://github.com/openai/math/blob/fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb/preprints/A-typical-start-upper-bound-for-low-temperature-SK-Glauber-dynamics-September-24-2026/paper.pdf. September 24, 2026. Source at commit fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb.

[86] Matteo Palassini. Ground-state energy fluctuations in the Sherrington–Kirkpatrick model. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10005, 2008. doi: 10.1088/1742-5468/2008/10/P10005.

[87] Vasyl Paliy and Youngtak Sohn. Rapid mixing within a phase for Ising models. https://arxiv.org/abs/2610.23777v1, 2026. URL https://arxiv.org/abs/2610.23777v1. Version 1, October 16, 2026.

[88] Dmitry Panchenko. *The Sherrington–Kirkpatrick Model*. Springer Monographs in Mathematics. Springer, New York, 2013. doi: 10.1007/978-1-4614-6289-7.

[89] Dmitry Panchenko. Free energy in the mixed p-spin models with vector spins. *The Annals of Probability*, 46(2):865–896, 2018. doi: 10.1214/17-AOP1194. URL https://doi.org/10.1214/17-AOP1194. https://doi.org/10.1214/17-AOP1194.

[90] Grigoris Paouris and Petros Valettas. A Gaussian small deviation inequality for convex functions. *The Annals of Probability*, 46(3):1441–1454, 2018. doi: 10.1214/17-AOP1206. URL https://arxiv.org/abs/1611.01723v2. Theorem and equation locators refer to arXiv:1611.01723v2, June 15, 2017.

[91] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754–1756, 1979. doi: 10.1103/PhysRevLett.43.1754. URL

https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.43.1754. https://doi.org/10.1103/PhysRevLett.43.1754.

[92] Giorgio Parisi and Tommaso Rizzo. Large deviations in the free energy of mean-field spin glasses. *Physical Review Letters*, 101:117205, 2008. doi: 10.1103/PhysRevLett.101.117205. URL https://arxiv.org/abs/0706.1180v2. Statement locators use https://arxiv.org/abs/0706.1180v2.

[93] Giorgio Parisi and Tommaso Rizzo. Phase diagram and large deviations in the free energy of mean-field spin glasses. *Physical Review B*, 79:134205, 2009. doi: 10.1103/PhysRevB.79.134205. URL https://arxiv.org/abs/0811.1524v2. Statement locators use https://arxiv.org/abs/0811.1524v2.

[94] Giorgio Parisi and Tommaso Rizzo. Universality and deviations in disordered systems. *Physical Review B*, 81:094201, 2010. doi: 10.1103/PhysRevB.81.094201. Equation locators refer to arXiv:0901.1100v1.

[95] Yan Ru Pei. Upper deviations of the Sherrington–Kirkpatrick free energy: the fifth-order law and the onset of constrained Parisi measures. https://arxiv.org/abs/2610.06480v2, 2026. URL https://arxiv.org/abs/2610.06480v2. Version 2, October 7, 2026.

[96] András Prékopa. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum (Szeged)*, 34:335–343, 1973. URL https://acta.bibl.u-szeged.hu/14411/. https://acta.bibl.u-szeged.hu/14411/1/math_034_335-343.pdf.

[97] David Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Communications in Mathematical Physics*, 108(2):225–239, 1987. doi: 10.1007/BF01210613.

[98] Adrien Schertzer. The order of free energy fluctuations in the critical Sherrington–Kirkpatrick model revisited. https://arxiv.org/abs/2606.21360v1, 2026. URL https://arxiv.org/abs/2606.21360v1. Version 1, June 19, 2026.

[99] Mark Sellke. Exponentially slow mixing of the low temperature SK model. https://arxiv.org/abs/2511.22621v2, 2025. URL https://arxiv.org/abs/2511.22621v2. Version 2, December 9, 2025.

[100] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792–1796, 1975. doi: 10.1103/PhysRevLett.35.1792. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.35.1792. https://doi.org/10.1103/PhysRevLett.35.1792.

[101] Haim Sompolinsky and Annette Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47:359–362, 1981. doi: 10.1103/PhysRevLett.47.359.

[102] Haim Sompolinsky and Annette Zippelius. Relaxational dynamics of the Edwards–Anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25:6860–6875, 1982. doi: 10.1103/PhysRevB.25.6860.

[103] Michel Talagrand. The Parisi formula. *Annals of Mathematics*, 163(1):221–263, 2006. doi: 10.4007/annals.2006.163.221. URL https://annals.math.princeton.edu/2006/163-1/p04. https://doi.org/10.4007/annals.2006.163.221.

[104] Michel Talagrand. Parisi measures. *Journal of Functional Analysis*, 231(2):269–286, 2006. doi: 10.1016/j.jfa.2005.03.001.

[105] Michel Talagrand. Mean field models for spin glasses: Some obnoxious problems. In Erwin Bolthausen and Anton Bovier, editors, *Spin Glasses*, volume 1900 of *Lecture Notes in Mathematics*, pages 63–80. Springer, Berlin, 2007. doi: 10.1007/978-3-540-40908-3_3. URL https://doi.org/10.1007/978-3-540-40908-3_3. https://doi.org/10.1007/978-3-540-40908-3_3.

[106] Michel Talagrand. Large deviations, Guerra's and A.S.S. schemes, and the Parisi hypothesis. *Journal of Statistical Physics*, 126(4-5):837–894, 2007. doi: 10.1007/s10955-006-9108-9.

[107] Michel Talagrand. *Mean Field Models for Spin Glasses. Volume I: Basic Examples*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Berlin, 2011. doi: 10.1007/978-3-642-15202-3.

[108] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 367–381. ACM, 2020. doi: 10.1145/3372885.3373824. URL https://doi.org/10.1145/3372885.3373824. https://doi.org/10.1145/3372885.3373824.

[109] D. J. Thouless, P. W. Anderson, and R. G. Palmer. Solution of 'Solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977. doi: 10.1080/14786437708235992.

[110] Miguel Tierz. Critical free-energy variance in the Sherrington–Kirkpatrick model. https://arxiv.org/abs/2610.00011v2, 2026. URL https://arxiv.org/abs/2610.00011v2. Version 2, October 16, 2026.

[111] Fabio Lucio Toninelli. About the Almeida–Thouless transition line in the Sherrington–Kirkpatrick mean-field spin glass model. *Europhysics Letters*, 60(5):764–767, 2002. doi: 10.1209/epl/i2002-00667-5.

[112] B. S. Tsirel'son, I. A. Ibragimov, and V. N. Sudakov. Norms of Gaussian sample functions. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory (Tashkent, 1975)*, volume 550 of *Lecture Notes in Mathematics*, pages 20–41. Springer, Berlin, 1976.

[113] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018. doi: 10.1017/9781108231596. URL https://webapps.math.uci.edu/~rvershyn/papers/HDP-book/HDP-1.pdf. First edition; numbered locators refer to this edition.

[114] D. Vertechi and M. A. Virasoro. Energy barriers in SK spin-glass model. *Journal de Physique*, 50(17):2325–2332, 1989. doi: 10.1051/jphys:0198900500170232500.

[115] Sihan Wang. Optimal mixing of Glauber dynamics for the Sherrington–Kirkpatrick model at $\beta < 1/2$. `https://arxiv.org/abs/2608.22159v2`, 2026. URL `https://arxiv.org/abs/2608.22159v2`. Version 2.

[116] Yuxin Zhou. Existence of full replica symmetry breaking for the Sherrington–Kirkpatrick model at low temperature. *Communications on Pure and Applied Mathematics*, 79(7):1746–1770, 2026. doi: 10.1002/cpa.70038.