The Annals of Probability
2019, Vol. 47, No. 3, 1587–1618
https://doi.org/10.1214/18-AOP1291

SUBOPTIMALITY OF LOCAL ALGORITHMS FOR A CLASS OF MAX-CUT PROBLEMS

BY WEI-KUO CHEN[*,1], DAVID GAMARNIK[‡], DMITRY PANCHENKO[†,2] AND MUSTAZEE RAHMAN[‡,2]

University of Minnesota[*], University of Toronto[†] and Massachusetts Institute of Technology[‡]

We show that in random K-uniform hypergraphs of constant average degree, for even $K \geq 4$, local algorithms defined as factors of i.i.d. can not find nearly maximal cuts, when the average degree is sufficiently large. These algorithms have been used frequently to obtain lower bounds for the max-cut problem on random graphs, but it was not known whether they could be successful in finding nearly maximal cuts. This result follows from the fact that the overlap of any two nearly maximal cuts in such hypergraphs does not take values in a certain nontrivial interval—a phenomenon referred to as the overlap gap property—which is proved by comparing diluted models with large average degree with appropriate fully connected spin glass models and showing the overlap gap property in the latter setting.

1. Introduction. This paper considers the problem of algorithmically finding nearly optimal spin configurations in the diluted K-spin model. We specifically focus on local algorithms defined as factors of i.i.d., the formal definition of which is provided in Section 2. The diluted K-spin model is also known as the max-cut problem for K-uniform Erdős–Rényi hypergraphs of constant average degree, and also as the random K-XORSAT model. The problem is only interesting for even K and we prove that, for even $K \geq 4$, local algorithms fail to find the nearly optimal spin configurations (maximal cuts) once the average degree is large enough.

The proof is based on finding a structural constraint for the overlap of any two nearly optimal spin configurations—the *overlap gap property*—that goes against certain properties of local algorithms. For $K = 2$, the overlap gap property is not expected to hold, which is why this case is excluded. The structural constraint is derived from recent results on the mean field K-spin spin glass models, in particular, the Parisi formula and the Guerra–Talagrand replica symmetry breaking bound at zero temperature. We begin with a discussion of the model and the notion of algorithms that we use.

Received July 2017; revised March 2018.

[1]Supported in part by NSF Grant DMS-1642207 and Hong Kong Research Grants Council GRF-14302515.

[2]Supported in part by NSERC.

MSC2010 subject classifications. Primary 60K35; secondary 05C80, 60G15, 60F10, 68W20, 82B44.

Key words and phrases. Local algorithms, maximum cut problems, spin glasses.

The K-spin model. The set of ± 1 spin configurations on N vertices will be denoted by

$$\Sigma^N = \{-1, 1\}^N.$$

Given any K-uniform hypergraph $G = ([N], E(G))$ on N vertices, a spin configuration $\sigma \in \Sigma^N$ and an edge $e = (v_1, \ldots, v_K) \in E(G)$, let us introduce the notation

$$\theta(\sigma, e) = -\sigma_{v_1} \cdots \sigma_{v_K}. \tag{1.1}$$

The K-spin Hamiltonian of G is the function $H_G : \Sigma^N \to \mathbb{R}$ defined by

$$H_G(\sigma) = \sum_{e \in E(G)} \theta(\sigma, e). \tag{1.2}$$

The *cut density* of σ is $H_G(\sigma)/N$. We consider directed hypergraphs and allow the same edge to appear multiple times, as well as a vertex to appear in an edge multiple times. However, in the case of the sparse Erdős–Rényi hypergraph that will be our main concern, this will happen only for an order one number of edges and vertices with high probability as the graph size tends to infinity.

The objective of the K-spin model for a given graph G is to maximize the cut density over all spin configurations and to find the maximizers or near maximizers. When edges consist of distinct vertices, this objective is trivial if K is odd with the maximum being achieved by the all (-1)-valued spin configuration. Therefore, throughout the paper we will assume that the parameter K is an even integer.

The diluted K-spin model. The K-uniform Erdős–Rényi hypergraph on N vertices with connectivity λ, denoted as $\mathrm{ER}(K, \lambda, N)$, is sampled as follows. First, sample a random number $|E| \sim \mathrm{Poisson}(\lambda N)$. Then for each $1 \le i \le |E|$, sample an *edge* $e_i \in [N]^K$ uniformly at random, where $[N] = \{1, \ldots, N\}$. The graph $\mathbf{G}_N \sim \mathrm{ER}(K, \lambda, N)$ has vertex set $V(\mathbf{G}_N) = [N]$ and edge set $E(\mathbf{G}_N) = \{e_1, \ldots, e_{|E|}\}$.

Let us consider the K-spin Hamiltonian for the graph $\mathrm{ER}(K, \lambda, N)$,

$$H_{\lambda,N}(\sigma) \triangleq H_{\mathbf{G}_N}(\sigma) = \sum_{e \in E(\mathbf{G})} \theta(\sigma, e). \tag{1.3}$$

This is the Hamiltonian of the so-called diluted K-spin model. Using Azuma's inequality, it is easy to show that the maximum of $H_{\lambda,N}(\sigma)/N$ over all $\sigma \in \Sigma^N$ is concentrated around its expectation as N increases. Moreover, the limit

$$M(K, \lambda) \triangleq \lim_{N \to \infty} \mathbb{E} \max_{\sigma \in \Sigma^N} \frac{H_{\lambda,N}(\sigma)}{N} \tag{1.4}$$

exists. This is proved for $K = 2$ in [7]. For general even $K \ge 2$, the existence of this limit follows from the same argument as in [13], which proves (for a different model) the existence of the thermodynamic limit of the free energy at positive temperature as a consequence of subadditivity obtained using an analogue of the

Guerra–Toninelli interpolation [23], which of course implies the existence of the limit at zero temperature. Computation of $M(K,\lambda)$ is challenging and, as we will discuss below, the appropriate large λ limit of $M(K,\lambda)$ recovers the famous Parisi formula, which expresses the ground state energy of the K-spin Sherrington–Kirkpatrick model.

Local algorithms. There is a considerable interest in devising algorithms that generate spin configurations with provably high cut density for general diluted models. Hatami et al. [25] conjectured that nearly optimal solutions to many such problems can be devised through local decision rules in the sense of factor of i.i.d. processes, as we explain below. Much effort has been made to address this question for problems such as perfect matchings (dimer covers), independent sets (hardcore model), max-cut (spin model), graph colouring (Potts model), etc.; see [6, 15, 16, 24, 26, 30, 31, 39] and references therein.

However, this conjecture was disproved in [19] for the independent set model by using an approach based on the overlap gap property for independent sets with sufficiently large cardinality, similar to the approach used in this paper. The tightest possible result for the same problem was obtained in [40] for half-optimal independent sets by looking at overlaps of many independents sets and establishing a multidimensional variant of the overlap gap property. The tightness of the result stems from the fact that less than half-optimal independent sets can be constructed by local algorithms [26]. Suboptimality of a sequential version of a local algorithm was established for the random NAE-K-SAT model in [20], also by establishing the overlap gap property for multi-overlaps. The conjecture has remained open for the diluted K-spin model, and, in this paper, we prove that for $K \geq 4$ and all large enough λ, local algorithms defined as factors of i.i.d. cannot find spin configurations whose cut density on $\mathrm{ER}(K,\lambda,N)$ is arbitrarily close to $M(K,\lambda)$.

Let us explain what we mean by local algorithms, specifically, algorithms defined as factors of i.i.d. A formal definition is provided in the next section. A *local algorithm* takes as input a graph G and decorates the vertices of G with i.i.d. random variables $X(v)$ for $v \in V(G)$. The algorithm then generates a spin configuration σ by deciding, for each vertex v, the value σ_v as a function (factor) of the randomly decorated r-neighborhood of v for some radius r that is independent of the size of G (hence the term "local"). The decision rule is the same for every vertex and is applied in parallel to all the vertices. The output of the algorithm is then a stochastic process over G.

Local algorithms are essentially randomized, distributed algorithms on graphs. They originate from ergodic theory where they are called factor of i.i.d. processes. Examples include continuous time Glauber dynamics run for a bounded time interval [5] and the Belief Propagation and Survey Propagation based algorithms run for a bounded number of iterations. Another well-known example is the Rödl nibble method for the packing problem. Interest in local algorithms arose partly because any stationary stochastic process over the Cayley graph of a finite generated

amenable group, such as $\mathbb{Z}^d$, can be approximated in distribution by local algorithms over the graph; see [30] for a discussion and references. On the other hand, there is a developing theory for processes that can be modeled by local algorithms over nonamenable graphs and expander graph sequences, such as random regular graphs or sparse Erdős–Rényi graphs with large enough connectivity (which behave like expanders in many ways due to the giant component). In this setting, local algorithms can generate various graph structures as previously indicated, and those constructions provide the best lower bounds to date for many optimization problems over sparse random graphs. Local algorithms can also approximate (in distribution) the local statistics of eigenvectors of random regular graphs [4, 24] and certain stationary Gaussian processes over transitive graphs [6].

Main result. Our main result about suboptimality of local algorithms is stated informally in Theorem 1 below and then precisely in Theorem 4, after appropriate terminology is introduced in Section 2. We now discuss the main result.

The *mean field*, or fully connected, K-spin Hamiltonian is the random function $H_N : \Sigma^N \to \mathbb{R}$ given by

$$H_N(\sigma) = \frac{1}{N^{(K-1)/2}} \sum_{(i_1,\dots,i_K)\in[N]^K} g_{i_1,\dots,i_K}\sigma_{i_1}\cdots\sigma_{i_K}, \tag{1.5}$$

where the $g_{i_1,\dots,i_K}$ are i.i.d. standard Gaussian random variables. Guerra and Toninelli proved in [22] that

$$\lim_{N\to\infty} \mathbb{E}\max_{\sigma\in\Sigma^N} \frac{H_N(\sigma)}{N} \triangleq P(K) \tag{1.6}$$

exists, and the precise description of this limit, given by the so called Parisi formula at zero temperature, will be discussed below and used crucially in the proof of the key results. Using another interpolation of Guerra and Toninelli from [23] that compares diluted models with large connectivity and fully connected models, Dembo et al. [14] (for $K = 2$) and Sen [41] showed that

$$M(K,\lambda) = P(K)\lambda^{1/2} + O\big(\lambda^{1/3}\big) \qquad \text{as } \lambda\to\infty. \tag{1.7}$$

In other words, the leading contribution to $M(K,\lambda)$ for large λ is $P(K)\lambda^{1/2}$. We show the following informal hardness of approximation result for $M(K,\lambda)$, which will be made more formal in Theorem 4 below.

THEOREM 1 (Informal). *For any even $K \geq 4$, there exists $\eta > 0$ such that the following holds for all large enough λ. Suppose A is a local algorithm constructed as a factor of i.i.d. that generates spin configurations on K-uniform hypergraphs and let $\sigma_A \in \Sigma^N$ be its output on the graph* $\mathrm{ER}(K,\lambda,N)$. *Then for all large enough N,*

$$\mathbb{E}\frac{H_{\lambda,N}(\sigma_A)}{N} \leq \big(P(K)-\eta\big)\lambda^{1/2}.$$

A summary of the proof. We informally describe the key steps in establishing this result. For $\eta > 0$, call a spin configuration $\sigma \in \Sigma^N$ to be η-optimal if the cut density $H_{\lambda,N}(\sigma)/N$ is at least $(P(K) - \eta)\lambda^{1/2}$. A key ingredient in the proof of the main result is the *overlap gap property*. It is the existence, when $K \geq 4$, of an $\eta_1 > 0$ and a positive-length interval $[a, b] \subset [0, 1]$ such that the absolute value of the overlap, defined in (1.9), of *any two* η_1-optimal spin configurations does not take values in $[a, b]$ with high probability (w.h.p.) as $N \to \infty$, provided that the connectivity λ is also sufficiently large. This is the content of Theorem 2.

For technical reasons, we need a strengthened version of the overlap gap property, namely, the overlap gap property for *coupled* diluted K-spin models. Specifically, we consider two coupled copies $\mathbf{G}^j_{N,t} \sim \mathrm{ER}(K, \lambda, N)$, $j = 1, 2$, where $t \in [0, 1]$ is a fixed parameter indicating some fraction of common edges in the two graphs. The two graphs are identical when $t = 1$ and are independent when $t = 0$. The details of the construction are given in Section 3.2. We establish the overlap gap property for this coupled model for the interval $[a, b]$. Namely, there exists an $\eta_2 > 0$ such that for all sufficiently large λ and for all $t \in [0, 1]$ the following holds: for every η_2-optimal solution σ_1 in $\mathbf{G}^1_{N,t}$ and every η_2-optimal solution σ_2 in $\mathbf{G}^2_{N,t}$, the absolute value of the overlap between σ_1 and σ_2 does not take value in $[a, b]$ w.h.p. as $N \to \infty$. This is the content of Theorem 5.

The overlap gap property for the coupled model discussed above is derived as a corollary of a similar overlap gap property, with the same interval $[a, b]$ as above, for the mean field K-spin model (1.5). The overlap gap property of the latter model is stated as Theorem 3 and the proof uses the Parisi formula at zero temperature for this model. The connection between the overlap gap property for the graph model of interest and the overlap gap property for the mean field model above is established via an interpolation technique that connects their respective nearly optimal spin configurations. The interpolation technique is also used in [12, 14, 27, 35, 41]. This is the content of Section 4.

Finally, the overlap gap property for the coupled model is used to establish limits on the largest cut achievable by local algorithms (defined as factors of i.i.d.). The proof is by contradiction, assuming that a local algorithm exists that provides a spin configuration achieving some η-optimal cut value in expectation. Standard concentration results can be used to show that this construction leads to an η-optimal cut w.h.p. as $N \to \infty$. Then we proceed by constructing two η-optimal and coupled spin configurations σ^1 and σ^2, based on factors of i.i.d., on the two t-coupled graphs $\mathbf{G}^1_{N,t}$ and $\mathbf{G}^2_{N,t}$, respectively. Here, we use t-coupled sources of i.i.d. decorations on the nodes of the coupled graph. In particular, when $t = 0$, the decorations are independent, and, since the graphs are independent as well, σ^1 and σ^2 are uncorrelated and have near zero asymptotical overlap. For this part, we also establish that every nearly optimal spin configuration is nearly balanced w.h.p.; see Lemma 3.3.

Next we show that the overlap of σ^1 and σ^2 is a continuous function of t. When $t = 1$, not only are the graphs identical but the spin configurations are as well since

they are based on the same set of i.i.d. decorations. Therefore, their overlap is 1 when $t = 1$. The continuity in t along with the endpoint values implies that for some t the overlap of σ^1 and σ^2 falls into the interval $[a, b]$. Thus, the cut values of σ^1 and σ^2 will not be η-optimal with the η referenced in Theorem 5. This is the contradiction.

Let us explain why we needed to establish the overlap gap property for the coupled model. If we proceeded with the same idea of t-coupled i.i.d. decorations on the *same* graph, we could not necessarily argue that the resulting spin configurations σ^1 and σ^2 have a small overlap when $t = 0$ since they would not be uncorrelated. In principle, they could have a nontrivial overlap due to the inherent randomness of neighborhood structures of a typical vertex in the $\mathrm{ER}(K, \lambda, N)$ model.

To conclude, we remark that the properties of local algorithms that were utilized in the proof of the main theorem are the following. First, concentration of the cut density and magnetization of spin configurations generated by any local algorithm f. Second, concentration of the overlap, $R_N(f, t)$, of the t-coupled spin configurations σ^1 and σ^2 generated from f via our construction. Third, the property that $t \to \mathbb{E}R_N(f, t)$ fills out a sufficiently dense subset of $[0, 1]$ for large N. Any class of algorithms that satisfy these properties will be suboptimal in the sense of Theorem 4 (or the informal Theorem 1).

1.1. *Overlap gap property.* The *magnetization* of a spin configuration $\sigma \in \Sigma^N$ is defined as

$$m(\sigma) = \frac{1}{N} \sum_{i \in [N]} \sigma_i. \tag{1.8}$$

The *overlap* of two spin configurations $\sigma^1, \sigma^2 \in \Sigma^N$ is defined as

$$R_{1,2} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^1 \sigma_i^2 = \frac{\sigma^1 \cdot \sigma^2}{N}. \tag{1.9}$$

As mentioned earlier, we need a statement of the aforementioned overlap gap property for coupled copies of the diluted K-spin model, which is Theorem 5 below. An already interesting illustration is the following special case. Here, $\mathbf{G}_N \sim \mathrm{ER}(K, \lambda, N)$ and $H_{\lambda,N}$ is its Hamiltonian as defined in (1.3).

THEOREM 2. *For any even $K \geq 4$, there exist $0 < a < b < 1$, $\eta_0 > 0$, $\lambda_0 > 0$ and $N_0 \geq 1$ such that, for $\lambda \geq \lambda_0$ and $N \geq N_0$, the following holds with probability at least $1 - Le^{-N/L}$ for some $L = L(\eta_0)$: for any two spin configurations σ^1, σ^2 that satisfy*

$$\frac{H_{\lambda,N}(\sigma^\ell)}{N} \geq \sqrt{\lambda}\big(P(K) - \eta_0\big), \qquad \ell = 1, 2,$$

the absolute value of their overlap satisfies $|R_{1,2}| \notin [a, b]$.

The proof of Theorem 2 (or more general Theorem 5 below) will based on approximating the diluted model in terms of the fully connected mean field model. The following theorem establishes the overlap gap property for the mean field K-spin Hamiltonian H_N in (1.5). Its proof is the most technical part of the paper.

THEOREM 3. *For even $K \geq 4$, there exist $0 < a < b < 1$ and $\eta > 0$ such that, for large enough N,*

$$\frac{1}{N}\mathbb{E} \max_{\substack{\sigma^1,\sigma^2 \in \Sigma^N: \\ |R_{1,2}| \in [a,b]}} \big(H_N(\sigma^1) + H_N(\sigma^2)\big) \leq \frac{2}{N}\mathbb{E} \max_{\sigma \in \Sigma^N} H_N(\sigma) - \eta. \tag{1.10}$$

The proof of Theorem 3 will utilize the zero temperature Parisi formula recently established by Auffinger and Chen in [2], and the Guerra–Talagrand replica symmetry breaking bound at zero temperature, which is an extension of the corresponding bound at positive temperature used earlier by Guerra and Talagrand [21, 42–44] to study the mean field K-spin model. Similar techniques have recently been used in [8] to study the questions of dynamics at positive temperature. For $K = 2$, the case of ordinary graphs, the overlap gap property (1.10) is not expected to hold. See [3] for a discussion and recent rigorous results on the overlap structure of mean field spin models.

Outline of the paper. Local algorithms are formally defined in Section 2 and the main result is stated in Theorem 4 following the necessary definitions. Section 3 contains the proof of Theorem 4. Section 4 contains the proof of Theorem 2 and Theorem 5, assuming the validity of Theorem 3. The proof of Theorem 3 is in Section 5.

2. Definitions and main result. We begin with some definitions. Formal statement of the main theorem, Theorem 4, is at the end of the section.

Hypergraphs and local algorithms. A K-uniform (directed) hypergraph is a pair $G = (V, E)$ where V is the set of vertices and $E \subseteq V^K$ are the edges, each edge being a K-tuple of vertices. We write $v \in e$ (v lies in e) if the edge e contains v as a coordinate. Two vertices are said to be connected if they lie in a common edge. The degree of a vertex is the number of edges that contain it, counted with multiplicity as a vertex may lie in an edge more than once. The distance between two vertices $v \neq w$ is the minimal integer $d \geq 1$ such that there exists a sequence of vertices $v_1, \ldots, v_{d+1}$ with $v = v_1$ and $w = v_{d+1}$ and with any two consecutive v_j, v_{j+1} for $1 \leq j \leq d$ belonging to the same edge. Distance from v to itself is zero. The r-neighborhood of v, denoted $B_r(G, v)$, is the subgraph of G induced by all vertices within distance r of v.

A rooted K-uniform hypergraph (G, v) is a K-uniform hypergraph with a distinguished vertex $v \in V(G)$ called the root. If r is the smallest integer such that $B_r(G, v) = G$ then we say (G, v) *has radius* r. Let

$$\mathcal{G}_r = \big\{(G, v) \mid (G, v) \text{ is a finite, connected, } K\text{-uniform rooted hypergraphs with radius at most } r\big\}. \tag{2.1}$$

Two rooted hypergraphs (G, v) and (G', v') are *isomorphic*, denoted $(G, v) \cong (G', v')$, if there is a bijection $\phi\colon V(G) \to V(G')$ such that $\phi(v) = v'$ and $e \in E(G)$ if and only if $\phi(e) \in E(G')$, where ϕ acts on edges coordinatewise. From now on, when we talk about hypergraphs, one can assume that we already picked a particular representative of an isomorphism class and think of isomorphisms as automorphisms of these representatives. In particular, we can redefine $\mathcal{G}_r$ as

$$\mathcal{G}_r = \big\{\text{representatives of isomorphism classes of (2.1)}\big\}. \tag{2.2}$$

For a sequence of random rooted hypergraphs (G_N, v_N), we define the sequence to *convergence in distribution* in a natural way as convergence of the probabilities $\mathbb{P}[B_r(G_N, v_N) \cong (G, v)]$ for every r and every $(G, v) \in \mathcal{G}_r$. In this case, some elementary measure theoretic and topological arguments imply that there is a random rooted graph (G_∞, v_∞) such that the aforementioned probabilities converge to $\mathbb{P}[B_r(G_\infty, v_\infty) \cong (G, v)]$. This notion of convergence is often called local weak convergence of graphs or also Benjamini–Schramm convergence; see [25] and references therein.

Factors on rooted, labeled hypergraphs. Let Λ_r denote the collection of all $[0, 1]$-*labeled hypergraphs* of radius at most r, that is, the collection of all triples (G, v, x), where:

1. $(G, v) \in \mathcal{G}_r$ defined in (2.2).
2. $x \in [0, 1]^{V(G)}$ is a labeling of vertices of G.

Two rooted $[0, 1]$-labeled hypergraphs (G, v, x) and (G', v', x') are *isomorphic* if there is a bijection $\phi\colon V(G) \to V(G')$, which is an isomorphism of (G, v) and (G', v'), and such that

$$x' = x \circ \phi^{-1}.$$

A function $f\colon \Lambda_r \to \{-1, 1\}$ is called a *factor of radius* r, denoted $f \in \mathcal{F}_r$, if it is measurable and

$$f(G, v, x) = f\big(G', v', x'\big)$$

for all isomorphic (G, v, x) and (G', v', x') in Λ_r. In other words, f is constant on isomorphism classes of Λ_r.

Local algorithms for K-uniform hypergraphs. A local algorithm A for K-uniform hypergraphs comes with an associated factor $f\colon \Lambda_r \to \{-1, 1\}$. The input to A is a finite K-uniform hypergraph G and a *random labeling* $X = (X(u))_{u\in V(G)}$ of $V(G)$ such that the $X(u)$'s are i.i.d. uniform random variables on $[0, 1]$. The *output* of the algorithm is a $\{-1, 1\}$-valued stochastic process $\sigma(f, X) = (\sigma(f, X)(u))_{u\in V(G)}$ defined by

$$\sigma(f, X)(u) = f\big(B_r(G, u), u, \big(X(w)\big)_{w\in B_r(G,u)}\big). \tag{2.3}$$

For simplicity of notation, when we write $w \in B_r(G, u)$ we really mean $w \in V(B_r(G, u))$. Notice that the radius of $B_r(G, u)$ can be smaller than r, which is why a factor f of radius r was defined on hypergraphs of radius less than or equal to r.

One obvious property of factors that we will use implicitly throughout the paper is the following. Suppose a factor $f \in \mathcal{F}_r$ outputs the spin configuration $\sigma(f, X)$ when run on a K-uniform hypergraph G. Then for any subgraph $G' \subseteq G$, the distribution of the restriction of $\sigma(X, f)$ onto G' depends only on the isomorphism class of the r-neighborhood of G' in G, that is, on the subgraph of G induced by $\bigcup_{v\in V(G')} B_r(G, v)$. Below, we will use the same property for coupled hypergraph models without mentioning it.

Main result. As above, $\mathbf{G}_N \overset{d}{=} \mathrm{ER}(K, \lambda, N)$ and $H_{\lambda,N}$ was defined in (1.3). Recall the definition of the quantity $P(K)$ from (1.6).

THEOREM 4. *Local algorithms defined as factors of i.i.d. cannot find spin configurations in* $\mathrm{ER}(K, \lambda, N)$ *that are nearly optimal in the following sense. There exists* $\eta > 0$ *such that*

$$\limsup_{\lambda\to\infty} \sup_{r} \sup_{f\in\mathcal{F}_r} \limsup_{N\to\infty} \mathbb{E}\frac{H_{\lambda,N}(\sigma(f, X))}{\sqrt{\lambda}N} \le P(K) - \eta. \tag{2.4}$$

The result establishes that, for large enough λ, the performance of any local algorithm is bounded away from the optimal value $P(K)$ by some additive gap η regardless of the choice of the factor f and its depth r. By the concentration result in Lemma 3.1 below, the same holds with high probability rather than on average.

3. Proof of the main result.

3.1. *Concentration of cut density and magnetization.* We begin by establishing a concentration of the cut density achievable by local algorithms.

LEMMA 3.1. *For every* λ, r, $f \in \mathcal{F}_r$ *and* $\epsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\big(\big|H_{\mathbf{G}_N}(\sigma(f, X)) - \mathbb{E}H_{\mathbf{G}_N}(\sigma(f, X))\big| \ge \epsilon N\big) = 0.$$

PROOF. For each node $v \in [N]$, let $N_\ell(v)$ denote the set of edges e containing v in the ℓth position of the edge, for $\ell = 1, 2, \ldots, K$. Let

$$H_{\mathbf{G}_N,v}(\sigma(f,X)) = \frac{1}{K}\sum_{1\le\ell\le K}\sum_{e\in N_\ell(v)}\theta(\sigma(f,X),e).$$

Observe that $H_{\mathbf{G}_N}(\sigma(f,X)) = \sum_{v\in[N]} H_{\mathbf{G}_N,v}(\sigma(f,X))$, since every edge e appears in exactly K terms. Let us expand the variance of $H_{\mathbf{G}_N}(\sigma(f,X))$ in terms of $H_{\mathbf{G}_N,v}(\sigma(f,X))$ for $v \in [N]$. We obtain

$$\begin{aligned}&\mathbb{E}\big(H_{\mathbf{G}_N}(\sigma(f,X)) - \mathbb{E}H_{\mathbf{G}_N}(\sigma(f,X))\big)^2\\ &\quad= \sum_{v,u\in[N]} \mathbb{E}\big(H_{\mathbf{G}_N,v}(\sigma(f,X)) - \mathbb{E}H_{\mathbf{G}_N,v}(\sigma(f,X))\big)\\ &\qquad\times\big(H_{\mathbf{G}_N,u}(\sigma(f,X)) - \mathbb{E}H_{\mathbf{G}_N,u}(\sigma(f,X))\big).\end{aligned}$$

Let us fix any two distinct vertices u and v in $[N]$, for instance, $u = 1$ and $v = 2$. A classical fact from the theory of random graph states that for every fixed r and λ, as $N \to \infty$, $B_{r+1}(\mathbf{G}_N, u) \cap B_{r+1}(\mathbf{G}_N, v) = \varnothing$ with high probability and the pair $(B_{r+1}(\mathbf{G}_N, u), B_{r+1}(\mathbf{G}_N, v))$ converges in distribution to a pair of independent Galton–Watson processes both truncated at $r+1$ generations. In this limiting Galton–Watson process (hypertree), for each vertex, the number of offsprings (i.e., hyperedges which contain this vertex and $K-1$ new distinct vertices that belong to the next generation) has Poisson(λK) distribution. Note that for every vertex v the value $H_{\mathbf{G}_N,v}(\sigma(f,X))$ is determined by $B_{r+1}(\mathbf{G}_N, v)$ and the i.i.d. labeling of the vertices in $B_{r+1}(\mathbf{G}_N, v)$. Therefore, we conclude that the random variables $H_{\mathbf{G}_N,v}(\sigma(f,X))$ and $H_{\mathbf{G}_N,u}(\sigma(f,X))$ are also asymptotically independent as $N \to \infty$. Both of these random variables are uniformly exponentially integrable in N, that is, for every $t > 0$, $\sup_N \mathbb{E}[\exp(t H_{\mathbf{G}_N,v}(\sigma(f,X)))] < \infty$. Thus,

$$\begin{aligned}&\lim_{N\to\infty}\mathbb{E}\big(H_{\mathbf{G}_N,v}(\sigma(f,X)) - \mathbb{E}H_{\mathbf{G}_N,v}(\sigma(f,X))\big)\\ &\qquad\times\big(H_{\mathbf{G}_N,u}(\sigma(f,X)) - \mathbb{E}H_{\mathbf{G}_N,u}(\sigma(f,X))\big) = 0.\end{aligned}$$

Notice that, by symmetry, this expectation is the same for all pairs of distinct vertices u and v, and we conclude that the variance of $H_{\mathbf{G}_N}(\sigma(f,X))$ satisfies

$$\mathbb{E}\big(H_{\mathbf{G}_N}(\sigma(f,X)) - \mathbb{E}H_{\mathbf{G}_N}(\sigma(f,X))\big)^2 = o(N^2)$$

as $N \to \infty$. Applying Chebyshev's inequality, we obtain the result. □

The identical proof also gives concentration of the magnetization.

LEMMA 3.2. *For every λ, r, $f \in \mathcal{F}_r$ and $\epsilon > 0$,*

$$\lim_{N\to\infty}\mathbb{P}\big(\big|m(\sigma(f,X)) - \mathbb{E}m(\sigma(f,X))\big| \ge \epsilon N\big) = 0.$$

Notice also that, by symmetry, $\mathbb{E}m(\sigma(f,X)) = \mathbb{E}\sigma_1(f,X)$, where σ_1 is the first coordinate of σ. In the proof of Theorem 4, we will need one more property of magnetization that we discuss next.

The proof of Theorem 4 below will proceed by contradiction, assuming that, for any $\eta > 0$, we can find arbitrarily large λ for which there exists a factor $f \in \mathcal{F}_r$ for some r such that

$$\mathbb{E}\frac{H_{\mathbf{G}_N}(\sigma(f,X))}{N} > \sqrt{\lambda}\big(P(K) - \eta\big) \tag{3.1}$$

for some arbitrarily large N. The concentration inequality in Lemma 3.1 then implies that, we can find arbitrarily large N for which, with high probability,

$$\frac{H_{\mathbf{G}_N}(\sigma(f,X))}{N} > \sqrt{\lambda}\big(P(K) - \eta\big). \tag{3.2}$$

It will be important to know that the magnetization $m(\sigma(f,X))$ of the output $\sigma(f,X)$ of the algorithm is small for large λ, with high probability. For this, as a byproduct of the interpolation technique in Section 4, we will prove the following.

LEMMA 3.3. *For any $\eta_0 > 0$, there exist λ_0 and N_0 such that for any $\lambda \geq \lambda_0$ and $N \geq N_0$, the following event holds with probability at least $1 - Le^{-N/L}$ for some $L = L(\eta)$: Whenever σ satisfies*

$$\frac{H_{\mathbf{G}_N}(\sigma)}{N} > \sqrt{\lambda}\big(P(K) - \eta_0\big),$$

it must also satisfy $|m(\sigma)| < \lambda^{-1/(2K)}(4\eta_0)^{1/K}$.

Take $\eta_0 = 1/8$ in the above. In (3.1) and (3.2), we can assume that $\eta \leq 1/8$, $\lambda \geq \lambda_0$ and $N \geq N_0$ and, as a result, with high probability we get that the absolute magnetization $|m(\sigma(f,X))| < \lambda^{-1/(2K)}2^{-1/K}$. Lemma 3.2 implies that $|\mathbb{E}m(\sigma(f,X))| \leq \lambda^{-1/(2K)}$, although the condition of how large N should be can now depend on f and λ. To summarize, we have the following.

LEMMA 3.4. *There exists λ_0 such that, for $\lambda \geq \lambda_0$, if (3.1) holds then $|\mathbb{E}m(\sigma(f,X))| \leq \lambda^{-1/(2K)}$, for all large enough N.*

3.2. *The coupled graph model and its properties.* The key idea in the proof of Theorem 4 will be running the local algorithm on a pair of coupled hypergraphs

$$\mathbf{G}^1_{N,t} \quad \text{and} \quad \mathbf{G}^2_{N,t} \tag{3.3}$$

defined on the same set of N vertices and indexed by the coupling parameter $t \in [0,1]$. The construction is as follows. Generate three independent random variables

$$|E_t| \stackrel{d}{=} \text{Poisson}(t\lambda N) \quad \text{and} \quad |E_{t,1}|, |E_{t,2}| \stackrel{d}{=} \text{Poisson}\big((1-t)\lambda N\big).$$

The somewhat unusual choice of notation $|E_\cdot|$ is motivated by having $E_\cdot$ denote sets of edges in graphs constructed below, where $|E_\cdot|$ are the cardinalities of these sets.

1. For each $1 \le i \le |E_t|$, generate an edge $e_i \in [N]^K$ uniformly at random, independently for all i. The set of these edges is denoted by E_t.
2. For each $\ell = 1, 2$ and each $1 \le i \le |E_{t,\ell}|$ generate an edge $e_i \in [N]^K$ uniformly at random, independently for all i, ℓ and independently from the set E_t. For a given ℓ, the set of these edges is denoted $E_{t,\ell}$.

The hypergraph $\mathbf{G}^\ell_{N,t}$ is defined by the edge set $E_t \cup E_{t,\ell}$ for $\ell = 1, 2$. Both $\mathbf{G}^1_{N,t}$ and $\mathbf{G}^2_{N,t}$ clearly have the same distribution $\mathrm{ER}(K, \lambda, N)$. When $t = 0$, the two hypergraphs are sampled independently of each other and, when $t = 1$, the two are identical.

The following overlap gap property for the coupled random graph model will be crucial in the proof of Theorem 4. It will be proved in Section 4. The special case corresponding to $t = 1$ was stated in Theorem 2 above.

THEOREM 5. *For any even $K \ge 4$, there exist $0 < a < b < 1$, $\eta_0 > 0$, $\lambda_0 > 0$ and $N_0 \ge 1$ such that, for any $t \in [0, 1]$, $\lambda \ge \lambda_0$, $N \ge N_0$, the following holds with probability at least $1 - Le^{-N/L}$ for some $L = L(\eta_0)$: whenever two spin configurations σ^1, σ^2 satisfy*

$$\frac{H_{\mathbf{G}^\ell_N}(\sigma^\ell)}{N} \ge \sqrt{\lambda}\big(P(K) - \eta_0\big)$$

then the absolute value of their overlap satisfies $|R_{1,2}| \notin [a, b]$.

To run local algorithms on the pair of hypergraphs $\mathbf{G}^1_{N,t}$ and $\mathbf{G}^2_{N,t}$ defined above, we need to define their labelings appropriately, as follows. Let us say that a vertex $v \in [N]$ is shared if it is incident to some shared edge in E_t, and nonshared otherwise. Consider a pair of independent random labelings X^1 and Y indexed by $[N]$. Define a new random labeling X^2 by

$$X^2(v) = \begin{cases} X^1(v) & \text{if } v \text{ is shared,} \\ Y(v) & \text{if } v \text{ is nonshared.} \end{cases}$$

Given a factor $f \in \mathcal{F}_r$, let $\sigma^\ell = \sigma(f, X^\ell)$ be its output on $\mathbf{G}^\ell_{N,t}$ for $\ell = 1, 2$ and let

$$R_N(f, t) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^1 \sigma_i^2 \tag{3.4}$$

be their overlap.

In the proof of Lemma 3.1, we used a classical fact that a neighborhood of a vertex in the Erdős–Rényi hypergraph converges to a Galton–Watson (hyper)tree

and, for any two distinct vertices, their neighborhood configurations are asymptotically independent. Of course, the same result holds for coupled hypergraphs, where in the limit we get coupled Galton–Watson processes defined as follows.

We start with a given vertex, call it a root, which is common to both trees. We generate Poisson$(t\lambda K)$ number of "shared" offsprings, where offspring of a vertex is a hyperedge containing this vertex and $K-1$ new vertices, and we also generate two independent Poisson$((1-t)\lambda K)$ sets of 'nonshared' offsprings for each tree. The new vertices in "shared" edges are labeled "shared" and new vertices in "nonshared" edges are labeled "nonshared." The root itself is labeled "nonshared" if Poisson$(t\lambda K)=0$, otherwise, it is labeled "shared" (this will be important for the labeling process below). From the second generation on, any "nonshared" vertex produces Poisson(λK) number of only "nonshared" edges, and any "shared" vertex produces Poisson$(t\lambda K)$ number of "shared" edges and two independent Poisson$((1-t)\lambda K)$ sets of "nonshared" edges, one for each tree.

Notice that the term "shared" does not correspond to the edges E_t in the construction $\mathbf{G}^1_{N,t}$ and $\mathbf{G}^2_{N,t}$ above, but rather has a meaning of "disjoint" in a local sense. Namely, for a given vertex v, its neighborhoods $B_r(\mathbf{G}^1_{N,t}, v)$ and $B_r(\mathbf{G}^2_{N,t}, v)$ can have edges belonging to E_t for any r. However, locally, for a finite r, the vertices in these two neighborhood are disjoint with high probability, unless they are all connected to v through edges in E_t. In the sense of local neighborhoods, in the limit "nonshared" corresponds to "disjoint" for finite N. In particular, since disjoint vertices are labeled by independent i.i.d. labelings, the labelings of two Galton–Watson trees will be the same for "shared" vertices and independent for "nonshared" vertices.

First of all, since for any two distinct vertices $u, v \in [N]$, their neighborhoods $(B_r(\mathbf{G}^1_{N,t}, v), B_r(\mathbf{G}^2_{N,t}, v))$ and $(B_r(\mathbf{G}^1_{N,t}, u), B_r(\mathbf{G}^2_{N,t}, u))$ are asymptotically independent labeled coupled Galton–Watson processes with r generation described above, the proof identical to Lemma 3.1 gives the following.

LEMMA 3.5. *For every λ, r, $f \in \mathcal{F}_r$, $t \in [0,1]$ and $\epsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\big(\big|R_N(f,t) - \mathbb{E}R_N(f,t)\big| \geq \epsilon\big) = 0.$$

Next, if we write, by symmetry, $\mathbb{E}R_N(f,t) = \mathbb{E}\sigma^1_1\sigma^2_1$ then the convergence of $(B_r(\mathbf{G}^1_{N,t}, 1), B_r(\mathbf{G}^2_{N,t}, 1))$ with its labelings to labeled coupled Galton–Watson process implies that

$$\lim_{N\to\infty} \mathbb{E}R_N(f,t) = \mathbb{E}\sigma^1\sigma^2 \triangleq R(f,t), \tag{3.5}$$

where σ^1 and σ^2 are outputs of the factor f at the root of these two Galton–Watson trees. Since there are countably many configurations for this coupled tree process and, obviously, the probability of each configuration is continuous in t, we have the following.

LEMMA 3.6. *The function $R(f,t)$ is continuous in t.*

Notice that $R(f,1)=1$, since for $t=1$ the two trees are completely "shared" with probability one, and $R(f,0)=(\mathbb{E}\sigma^1)^2$, since for $t=0$ the trees are completely independent, or "nonshared," with probability one. Moreover, $\mathbb{E}\sigma^1$ is the limit of magnetization,

$$\lim_{N\to\infty}\mathbb{E}m\big(\sigma(f,X)\big)=\lim_{N\to\infty}\mathbb{E}\sigma_1(f,X)=\mathbb{E}\sigma^1. \tag{3.6}$$

With this we are ready to prove Theorem 4.

3.3. *Completing the proof of Theorem 4.* Consider a, b, η_0 defined in Theorem 5 and let λ_0 be the larger of the two values λ_0 defined in Theorem 5 and Lemma 3.4.

The proof of Theorem 4 below will proceed by contradiction. Assume that, for $\eta=\eta_0/2$, we can find

$$\lambda\ge\max\big(\lambda_0,a^{-K}\big),$$

for which there exists a factor $f\in\mathcal{F}_r$ for some r such that

$$\mathbb{E}\frac{H_{\mathbf{G}_N}(\sigma(f,X))}{N}>\sqrt{\lambda}\big(P(K)-\eta\big) \tag{3.7}$$

for arbitrarily large N. Lemma 3.4 together with (3.6) implies that $|\mathbb{E}\sigma^1|\le\lambda^{-1/(2K)}$ and, therefore,

$$R(f,0)=\big(\mathbb{E}\sigma^1\big)^2\le\lambda^{-1/K}<a.$$

Since $R(f,t)$ in (3.5) is continuous, $R(f,0)<a$ and $R(f,1)=1$, there exists $t\in(0,1)$ such that $R(f,t)=(a+b)/2$. For this t, (3.5) and Lemma 3.5 imply that, with high probability, the overlap $R_N(f,t)$ will belong to the interval (a,b) for large enough N. On the other hand, (3.7) and Lemma 3.1 imply that

$$\frac{H_{\mathbf{G}_N^\ell}(\sigma^\ell)}{N}\ge\sqrt{\lambda}\big(P(K)-\eta_0\big),$$

with probability going to one as $N\to\infty$. By Theorem 5, with high probability we must have that $R_N(f,t)$ does not belong to the interval $[a,b]$, which is a contradiction.

4. Overlap gap in the diluted model. This section proves Theorem 5, as well as Theorem 2 as a simpler illustration. In order to prove these results, we will first use the Guerra–Toninelli interpolation from [23] to connect the diluted K-spin model to the mean field K-spin model. This connection has been used several times in the literature [12, 14, 27, 35, 41]. Theorem 2 will then be deduced from the conclusion of Theorem 3.

Before starting with the proof, we set some notation that will be used throughout the remaining Sections 4 and 5. For $\lambda > 0$, let H_λ denote the Hamiltonian of the diluted K-spin model with connectivity λ on N vertices, that is, the K-spin Hamiltonian for the hypergraph $\mathrm{ER}(K, \lambda, N)$. Recall the Hamiltonians $H_{\mathbf{G}^1_{N,t}}$ and $H_{\mathbf{G}^2_{N,t}}$ from (3.3). Denote them respectively by H^1_λ and H^2_λ (we keep the dependence on the coupling parameter t implicit). We also denote by H the Hamiltonian of the mean field, or fully connected, K-spin model on N vertices as defined in (1.5). Let H' and H'' be independent copies of H. For $t \in [0, 1]$, set

$$\begin{aligned} H^1 &= \sqrt{t}H + \sqrt{1-t}H', \\ H^2 &= \sqrt{t}H + \sqrt{1-t}H''. \end{aligned} \tag{4.1}$$

We will assume throughout that N is even and suppress the dependence on t and N for clarity.

First we establish a key interpolation lemma. Let $A \subseteq \Sigma^N$, $S \subseteq [0, 1]$, $t \in [0, 1]$ and $\lambda > 0$ be fixed. For $s \in [0, 1]$, consider the interpolating Hamiltonian

$$H(s, \sigma^1, \sigma^2) = \sum_{\ell=1}^{2} \bigl(\delta H^\ell_{\lambda(1-s)}(\sigma^\ell) + \sqrt{s}\beta H^\ell(\sigma^\ell)\bigr), \tag{4.2}$$

where the Hamiltonians $H^\ell_{\lambda(1-s)}$ correspond to the coupled diluted K-spin model with the coupling parameter t and connectivity parameter $\lambda(1-s)$. The parameters $\delta > 0$ and $\beta > 0$ will be chosen later. Let

$$\varphi(s) = \frac{1}{N}\mathbb{E}\log \sum_{\sigma^1,\sigma^2 \in A : |R_{1,2}| \in S} \exp H(s, \sigma^1, \sigma^2). \tag{4.3}$$

Let us denote by $\langle\cdot\rangle_s$ the average with respect to the Gibbs' measure

$$G_s(\sigma^1, \sigma^2) = \frac{\exp H(s, \sigma^1, \sigma^2)}{\sum_{\sigma^1,\sigma^2 \in A : |R_{1,2}| \in S} \exp H(s, \sigma^1, \sigma^2)} \tag{4.4}$$

defined on $\{(\sigma^1, \sigma^2) \in A \times A : |R_{1,2}| \in S\}$. The following holds.

LEMMA 4.1. *For any $A \subseteq \Sigma^N$, $S \subseteq [0, 1]$, $t \in [0, 1]$ and $\lambda > 0$,*

$$\begin{aligned} &\frac{1}{N}\mathbb{E} \max_{\sigma^1,\sigma^2 \in A : |R_{1,2}| \in S} \bigl(H^1_\lambda(\sigma^1) + H^2_\lambda(\sigma^2)\bigr) \\ &\quad = \frac{\sqrt{\lambda}}{N}\mathbb{E} \max_{\sigma^1,\sigma^2 \in A : |R_{1,2}| \in S} \bigl(H^1(\sigma^1) + H^2(\sigma^2)\bigr) \\ &\qquad - \lambda \int_0^1 \mathbb{E}\bigl\langle m(\sigma^1)^K + m(\sigma^2)^K \bigr\rangle_s \, ds + O(\lambda^{1/3}), \end{aligned} \tag{4.5}$$

where $O(\lambda^{1/3}) \le L\lambda^{1/3}$ for some $L > 0$ independent of N, A, S, t, λ.

PROOF. The proof will make use of the interpolating Hamiltonian (4.2) with appropriate choices for β and δ in terms of λ. First let us compute the derivative $\varphi'(s) = \mathrm{I} + \mathrm{II}$ as a sum of two terms coming from the Gaussian integration by parts and Poisson integration by parts. Let us denote the i.i.d. samples (replicas) from G_s by $(\sigma^{j,1}, \sigma^{j,2})$ for $j \geq 1$, and denote the overlap

$$R_{\ell,\ell'}^{j,j'} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^{j,\ell}\sigma_i^{j',\ell'}.$$

Taking the derivative in $\sqrt{s}$ and using Gaussian integration by parts (see, e.g., [32], Section 1.2) gives

$$\mathrm{I} = \frac{\beta^2}{2}\bigl(2 + 2t\mathbb{E}\bigl\langle \bigl(R_{1,2}^{1,1}\bigr)^K\bigr\rangle_s - \mathbb{E}\bigl\langle \bigl(R_{1,1}^{1,2}\bigr)^K + \bigl(R_{2,2}^{1,2}\bigr)^K + t\bigl(R_{1,2}^{1,2}\bigr)^K + t\bigl(R_{2,1}^{1,2}\bigr)^K\bigr\rangle_s\bigr).$$

The rest of the calculation below is as in the proof of the Franz–Leone upper bound for the free energy in diluted models in [18, 36]. In order to differentiate $\varphi(s)$ with respect to the Poisson random variables involved, we use that

$$\frac{d}{ds}\mathbb{E}f\bigl(\mathrm{Poisson}(s)\bigr) = \mathbb{E}f\bigl(\mathrm{Poisson}(s)+1\bigr) - \mathbb{E}f\bigl(\mathrm{Poisson}(s)\bigr)$$

for a Poisson random variable $\mathrm{Poisson}(s)$ having mean s. If a function depends on several independent Poisson random variables, by the product rule, the derivative will be a sum of such differences, where $+1$ is added to one Poisson random variable at a time. Hence, with the notation $\sum := \sum_{\sigma^1,\sigma^2 \in A : |R_{1,2}| \in S}$,

$$\begin{aligned}\mathrm{II} = {} & -t\lambda\Bigl(\mathbb{E}\log\sum \exp H^+\bigl(s,\sigma^1,\sigma^2\bigr) - \mathbb{E}\log\sum\exp H\bigl(s,\sigma^1,\sigma^2\bigr)\Bigr)\\ & - (1-t)\lambda\Bigl(\mathbb{E}\log\sum \exp H_1^+\bigl(s,\sigma^1,\sigma^2\bigr) - \mathbb{E}\log\sum\exp H\bigl(s,\sigma^1,\sigma^2\bigr)\Bigr)\\ & - (1-t)\lambda\Bigl(\mathbb{E}\log\sum \exp H_2^+\bigl(s,\sigma^1,\sigma^2\bigr) - \mathbb{E}\log\sum\exp H\bigl(s,\sigma^1,\sigma^2\bigr)\Bigr),\end{aligned}$$

where $H^+(s,\sigma^1,\sigma^2)$ includes one additional common edge $(i_1,\dots,i_K)$ for both configurations σ^1,σ^2 (this new edge is chosen independently and uniformly at random), and $H_1^+(s,\sigma^1,\sigma^2)$ and $H_2^+(s,\sigma^1,\sigma^2)$ include an additional edge on either the first or second configurations, respectively. That is,

$$\begin{aligned}H^+\bigl(s,\sigma^1,\sigma^2\bigr) &= H\bigl(s,\sigma^1,\sigma^2\bigr) - \delta\sigma_{i_1}^1\cdots\sigma_{i_K}^1 - \delta\sigma_{i_1}^2\cdots\sigma_{i_K}^2,\\ H_1^+\bigl(s,\sigma^1,\sigma^2\bigr) &= H\bigl(s,\sigma^1,\sigma^2\bigr) - \delta\sigma_{i_1}^1\cdots\sigma_{i_K}^1,\\ H_2^+\bigl(s,\sigma^1,\sigma^2\bigr) &= H\bigl(s,\sigma^1,\sigma^2\bigr) - \delta\sigma_{i_1}^2\cdots\sigma_{i_K}^2.\end{aligned}$$

We can then rewrite the above as

$$\begin{aligned}\mathrm{II} = {} & -t\lambda\mathbb{E}\log\bigl\langle\exp\bigl(-\delta\sigma_{i_1}^1\cdots\sigma_{i_K}^1 - \delta\sigma_{i_1}^2\cdots\sigma_{i_K}^2\bigr)\bigr\rangle_s\\ & - (1-t)\lambda\mathbb{E}\log\bigl\langle\exp\bigl(-\delta\sigma_{i_1}^1\cdots\sigma_{i_K}^1\bigr)\bigr\rangle_s - (1-t)\lambda\mathbb{E}\log\bigl\langle\exp\bigl(-\delta\sigma_{i_1}^2\cdots\sigma_{i_K}^2\bigr)\bigr\rangle_s.\end{aligned}$$

Since the product of spins takes values ± 1, we can represent

$$\exp\bigl(-\delta\sigma_{i_1}^{\ell}\cdots\sigma_{i_K}^{\ell}\bigr)=\operatorname{ch}(\delta)\bigl(1-\operatorname{th}(\delta)\sigma_{i_1}^{\ell}\cdots\sigma_{i_K}^{\ell}\bigr)$$

and

$$\exp\bigl(-\delta\sigma_{i_1}^{1}\cdots\sigma_{i_K}^{1}-\delta\sigma_{i_1}^{2}\cdots\sigma_{i_K}^{2}\bigr)=\operatorname{ch}(\delta)^2\bigl(1-\operatorname{th}(\delta)\Delta\bigl(\sigma^1,\sigma^2\bigr)\bigr)$$

with the notation

$$\Delta\bigl(\sigma^1,\sigma^2\bigr)=\sigma_{i_1}^{1}\cdots\sigma_{i_K}^{1}+\sigma_{i_1}^{2}\cdots\sigma_{i_K}^{2}-\operatorname{th}(\delta)\sigma_{i_1}^{1}\cdots\sigma_{i_K}^{1}\sigma_{i_1}^{2}\cdots\sigma_{i_K}^{2},$$

where ch() and th() are hyperbolic cosine and tangent, respectively. Expressing the logarithm by its Taylor series, we can rewrite

$$\mathrm{II}=-2\lambda\log\operatorname{ch}\delta+\lambda\sum_{n\geq 1}\frac{\operatorname{th}(\delta)^n}{n}\biggl(t\mathbb{E}\bigl\langle\Delta\bigl(\sigma^1,\sigma^2\bigr)\bigr\rangle_s^n+(1-t)\sum_{\ell=1}^{2}\mathbb{E}\bigl\langle\sigma_{i_1}^{\ell}\cdots\sigma_{i_K}^{\ell}\bigr\rangle_s^n\biggr).$$

Using replicas, we can represent

$$\mathbb{E}\bigl\langle\Delta\bigl(\sigma^1,\sigma^2\bigr)\bigr\rangle_s^n=\mathbb{E}\biggl\langle\prod_{j\leq n}\Delta\bigl(\sigma^{j,1},\sigma^{j,2}\bigr)\biggr\rangle_s=\mathbb{E}\biggl\langle\mathbb{E}'\prod_{j\leq n}\Delta\bigl(\sigma^{j,1},\sigma^{j,2}\bigr)\biggr\rangle_s,$$

where $\mathbb{E}'$ is the expectation with respect to the random indices $i_1,\dots,i_K$, which are independent of the random variables in $\langle\cdot\rangle_s$. For $n=1$,

$$\mathbb{E}'\sigma_{i_1}^{\ell}\cdots\sigma_{i_K}^{\ell}=m\bigl(\sigma^{\ell}\bigr)^K$$

and

$$\mathbb{E}'\Delta\bigl(\sigma^{1,1},\sigma^{1,2}\bigr)=m\bigl(\sigma^{1,1}\bigr)^K+m\bigl(\sigma^{1,2}\bigr)^K-\operatorname{th}(\delta)\bigl(R_{1,2}^{1,1}\bigr)^K.$$

Therefore,

$$\begin{aligned}
&t\mathbb{E}\bigl\langle\Delta\bigl(\sigma^1,\sigma^2\bigr)\bigr\rangle_s+(1-t)\mathbb{E}\bigl\langle\sigma_{i_1}^{1}\cdots\sigma_{i_K}^{1}\bigr\rangle_s+(1-t)\mathbb{E}\bigl\langle\sigma_{i_1}^{2}\cdots\sigma_{i_K}^{2}\bigr\rangle_s\\
&\quad=t\bigl(\mathbb{E}\bigl\langle m\bigl(\sigma^{1,1}\bigr)^K+m\bigl(\sigma^{1,2}\bigr)^K\bigr\rangle_s-\operatorname{th}(\delta)\mathbb{E}\bigl\langle\bigl(R_{1,2}^{1,1}\bigr)^K\bigr\rangle_s\bigr)\\
&\qquad+(1-t)\mathbb{E}\bigl\langle m\bigl(\sigma^{1,1}\bigr)^K+m\bigl(\sigma^{1,2}\bigr)^K\bigr\rangle_s\\
&\quad=\mathbb{E}\bigl\langle m\bigl(\sigma^{1,1}\bigr)^K+m\bigl(\sigma^{1,2}\bigr)^K\bigr\rangle_s-t\operatorname{th}(\delta)\mathbb{E}\bigl\langle\bigl(R_{1,2}^{1,1}\bigr)^K\bigr\rangle_s.
\end{aligned}$$

For $n=2$, we compute similarly

$$\mathbb{E}\bigl\langle\Delta\bigl(\sigma^1,\sigma^2\bigr)\bigr\rangle_s^2=\mathbb{E}\bigl\langle\bigl(R_{1,1}^{1,2}\bigr)^K+\bigl(R_{2,2}^{1,2}\bigr)^K+\bigl(R_{1,2}^{1,2}\bigr)^K+\bigl(R_{2,1}^{1,2}\bigr)^K\bigr\rangle_s+\mathrm{III}_1$$

and

$$\sum_{\ell=1}^{2}\mathbb{E}\bigl\langle\sigma_{i_1}^{\ell}\cdots\sigma_{i_K}^{\ell}\bigr\rangle_s^2=\mathbb{E}\bigl\langle\bigl(R_{1,1}^{1,2}\bigr)^K+\bigl(R_{2,2}^{1,2}\bigr)^K\bigr\rangle_s,$$

where $|\mathrm{III}_1| \le L\,\mathrm{th}(\delta)$. Finally, since $|\Delta(\sigma^1,\sigma^2)| \le 2$ and $|\sigma^\ell_{i_1}\cdots\sigma^\ell_{i_K}| = 1$, we can bound

$$\left| t\lambda \sum_{n\ge 3} \frac{\mathrm{th}(\delta)^n}{n} \mathbb{E}\langle \Delta(\sigma^1,\sigma^2)\rangle_s^n \right| \le \lambda \sum_{n\ge 3} \frac{(2\,\mathrm{th}(\delta))^n}{n} \le L\lambda\delta^3$$

and

$$\left| (1-t)\lambda \sum_{n\ge 3} \frac{\mathrm{th}(\delta)^n}{n} \sum_{\ell=1}^{2} \mathbb{E}\langle \sigma^\ell_{i_1}\cdots\sigma^\ell_{i_K}\rangle_s^n \right| \le L\lambda\delta^3$$

for small enough $\delta > 0$, where L is a universal constant independent of everything. Collecting all the terms,

$$\begin{aligned}\mathrm{II} = &-2\lambda\log\mathrm{ch}\,\delta + \lambda\,\mathrm{th}(\delta)\mathbb{E}\langle m(\sigma^{1,1})^K + m(\sigma^{1,2})^K\rangle_s - t\lambda\,\mathrm{th}(\delta)^2\mathbb{E}\langle (R^{1,1}_{1,2})^K\rangle_s \\ &+ \frac{\lambda\,\mathrm{th}(\delta)^2}{2}\mathbb{E}\langle (R^{1,2}_{1,1})^K + (R^{1,2}_{2,2})^K + t(R^{1,2}_{1,2})^K + t(R^{1,2}_{2,1})^K\rangle_s + O(\lambda\delta^3).\end{aligned}$$

Next, for a given λ and δ, we will define β by

$$\beta = \sqrt{\lambda}\,\mathrm{th}(\delta). \tag{4.6}$$

With this choice, all the terms in $\varphi'(s) = \mathrm{I} + \mathrm{II}$ containing the overlaps $R^{j,j'}_{\ell,\ell'}$ cancel out and we get

$$\begin{aligned}\varphi'(s) &= \mathrm{I} + \mathrm{II} \\ &= -2\lambda\log\mathrm{ch}\,\delta + \lambda\,\mathrm{th}(\delta)\mathbb{E}\langle m(\sigma^{1,1})^K + m(\sigma^{1,2})^K\rangle_s + \lambda\,\mathrm{th}(\delta)^2 + O(\lambda\delta^3).\end{aligned}$$

One can check that $-2\log\mathrm{ch}\,\delta + \mathrm{th}(\delta)^2 = O(\delta^4)$ as $\delta \to 0$ and, therefore,

$$\varphi'(s) = \lambda\delta\mathbb{E}\langle m(\sigma^{1,1})^K + m(\sigma^{1,2})^K\rangle_s + O(\lambda\delta^3).$$

Now, integrating between 0 and 1 and dividing both sides by δ,

$$\begin{aligned}\frac{1}{\delta}\varphi(1) - \frac{1}{\delta}\varphi(0) &= \frac{1}{\delta}\int_0^1 \phi'(s)\,ds \\ &= \lambda\int_0^1 \mathbb{E}\langle m(\sigma^{1,1})^K + m(\sigma^{1,2})^K\rangle_s\,ds + O(\lambda\delta^2).\end{aligned}$$

Using the elementary estimates,

$$\begin{aligned}\frac{1}{N}\mathbb{E}\max_{\sigma^1,\sigma^2\in A:|R_{1,2}|\in S} H(s,\sigma^1,\sigma^2) &\le \varphi(s) \\ &\le 2\log 2 + \frac{1}{N}\mathbb{E}\max_{\sigma^1,\sigma^2\in A:|R_{1,2}|\in S} H(s,\sigma^1,\sigma^2)\end{aligned}$$

at $s = 0$ and $s = 1$, we infer from the above that

$$\left|\frac{1}{N}\mathbb{E}\max_{\sigma^1,\sigma^2\in A:|R_{1,2}|\in S}\bigl(H_\lambda^1(\sigma^1)+H_\lambda^2(\sigma^2)\bigr)-\frac{1}{\delta}\varphi(0)\right|\le\frac{2\log 2}{\delta},$$

$$\left|\frac{\beta}{\delta}\frac{1}{N}\mathbb{E}\max_{\sigma^1,\sigma^2\in A:|R_{1,2}|\in S}\bigl(H^1(\sigma^1)+H^2(\sigma^2)\bigr)-\frac{1}{\delta}\varphi(1)\right|\le\frac{2\log 2}{\delta}.$$

By Taylor's expansion and our choice of β,

$$\frac{\beta}{\delta}=\frac{\sqrt{\lambda}\,\mathrm{th}(\delta)}{\delta}=\sqrt{\lambda}+O\bigl(\sqrt{\lambda}\delta^2\bigr)$$

and, therefore,

$$\begin{aligned}&\frac{1}{N}\mathbb{E}\max_{\sigma^1,\sigma^2\in A:|R_{1,2}|\in S}\bigl(H_\lambda^1(\sigma^1)+H_\lambda^2(\sigma^2)\bigr)\\&\quad=\frac{\sqrt{\lambda}}{N}\mathbb{E}\max_{\sigma^1,\sigma^2\in A:|R_{1,2}|\in S}\bigl(H^1(\sigma^1)+H^2(\sigma^2)\bigr)\\&\qquad-\lambda\int_0^1\mathbb{E}\bigl\langle m\bigl(\sigma^{1,1}\bigr)^K+m\bigl(\sigma^{1,2}\bigr)^K\bigr\rangle_s\,ds+O\biggl(\frac{1}{\delta}+\lambda\delta^2+\sqrt{\lambda}\delta^2\biggr).\end{aligned}$$

Letting $\delta=\lambda^{-1/3}$ in the last equation completes our proof. □

Using the above interpolation and Theorem 3 (which will be proved in the following section), we are now ready to prove Theorem 5.

PROOF OF THEOREM 5. Set $S=[a,b]$ as in Theorem 3 and define

$$f(t):=\limsup_{N\to\infty}\frac{1}{N}\mathbb{E}\max_{|R_{1,2}|\in S}\bigl(H^1(\sigma^1)+H^2(\sigma^2)\bigr),$$

where H^1, H^2 were defined in (4.1). It is easy to see that f is a continuous function of t. By Theorem 3 above and [12], Theorem 2 (used with $\varepsilon=a/2$ there), we have that

$$f(t)<2P(K)\qquad\text{for all } t\in[0,1].$$

[*Remark*: In fact, the case of K-spin model was established in Theorem 2 in [11] and stated for $t\in(0,1)$ in both [11, 12]. However, the case of $t=0$ is implicitly included in the proof of Theorem 2 in [11] and is in fact the easiest case. One should simply note that, in the notation of that paper, the derivative of the Guerra–Talagrand upper bound for $t=0$ with respect to the Lagrange multiplier λ equals minus the overlap constraint $-q$, which results in energy penalty for nonzero overlap constraints.]

The continuity of f ensures that there exists some $\eta_0>0$ such that $f(t)<2P(K)-6\eta_0$ for all $t\in[0,1]$. Using Lemma 4.1 with $A=\Sigma^N$ and noting that

the terms involving magnetization are negative, we see that, for any $t \in [0,1]$ and large enough $N \geq N_0$,

$$\frac{1}{N}\mathbb{E}\max_{|R_{1,2}|\in S}\bigl(H_\lambda^1(\sigma^1)+H_\lambda^2(\sigma^2)\bigr) \leq \frac{\sqrt{\lambda}}{N}\mathbb{E}\max_{|R_{1,2}|\in S}\bigl(H^1(\sigma^1)+H^2(\sigma^2)\bigr)+L\lambda^{1/3}$$
$$\leq \sqrt{\lambda}\bigl(2P(K)-6\eta_0\bigr)+L\lambda^{1/3}$$

and, thus, if $\lambda \geq \lambda_0 = (L\eta_0)^{-6}$ for some large enough L then

$$\frac{1}{N}\mathbb{E}\max_{\substack{\sigma^1,\sigma^2\in A:\\ |R_{1,2}|\in S}}\bigl(H_\lambda^1(\sigma^1)+H_\lambda^2(\sigma^2)\bigr) \leq 2\sqrt{\lambda}\bigl(P(K)-2\eta_0\bigr).$$

By a standard concentration inequality for Poisson(λN) and Azuma's concentration inequality, one can check that there exists $L = L(\eta_0)$ such that, with probability at least $1 - Le^{-N/L}$,

$$\frac{1}{N}\max_{|R_{1,2}|\in S}\bigl(H_\lambda^1(\sigma^1)+H_\lambda^2(\sigma^2)\bigr)$$
$$\leq \frac{1}{N}\mathbb{E}\max_{|R_{1,2}|\in S}\bigl(H_\lambda^1(\sigma^1)+H_\lambda^2(\sigma^2)\bigr)+\eta_0\sqrt{\lambda}$$
$$\leq 2\sqrt{\lambda}\bigl(P(K)-\eta_0\bigr).$$

On this event, if σ^1, σ^2 satisfy

$$\frac{H_\lambda^\ell(\sigma^\ell)}{N} > \sqrt{\lambda}\bigl(P(K)-\eta_0\bigr), \qquad \ell = 1,2,$$

then $|R(\sigma^1,\sigma^2)| \notin S = [a,b]$, which finishes the proof. □

Finally, we use the interpolation to prove Lemma 3.3, which shows that near maximizers automatically have small magnetization.

PROOF OF LEMMA 3.3. Let $c = (4\eta_0)^{1/K}\lambda^{-1/(2K)}$. In the setting of Lemma 4.1, let $A = \{\sigma \in \Sigma^N : |m(\sigma)| \geq c\}$, $S = [0,1]$ and $t = 1$. In this case, $H_\lambda^1 = H_\lambda^2 = H_\lambda$ and $H^1 = H^2 = H$. Since

$$\lambda\int_0^1 \mathbb{E}\bigl\langle m\bigl(\sigma^1\bigr)^K + m\bigl(\sigma^2\bigr)^K\bigr\rangle_s\,ds \geq 2\lambda c^K = 8\sqrt{\lambda}\eta_0,$$

we have

$$\frac{1}{N}\mathbb{E}\max_{\sigma\in A}H_\lambda(\sigma) \leq \frac{\sqrt{\lambda}}{N}\mathbb{E}\max_{\sigma\in A}H(\sigma) - 4\sqrt{\lambda}\eta_0 + L\lambda^{1/3}$$

and, thus, there exists N_0 such that, for all $N \geq N_0$,

$$\frac{1}{N}\mathbb{E}\max_{\sigma\in A} H_\lambda(\sigma) \leq \frac{\sqrt{\lambda}}{N}\mathbb{E}\max_{\sigma\in\Sigma^N} H(\sigma) - 4\sqrt{\lambda}\eta_0 + O\big(\lambda^{1/3}\big)$$
$$\leq \sqrt{\lambda}\big(P(K) - 3\eta_0\big) + L\lambda^{1/3}.$$

Consequently, for $N \geq N_0$ and $\lambda \geq L^6\eta_0^{-6}$,

$$\frac{1}{N}\mathbb{E}\max_{\sigma\in A} H_\lambda(\sigma) \leq \sqrt{\lambda}\big(P(K) - 2\eta_0\big).$$

From standard concentration inequality for Poisson(λN) and Azuma's concentration inequality, one can check that there exists $L = L(\eta_0)$ such that, with probability at least $1 - Le^{-N/L}$,

$$\frac{1}{N}\max_{\sigma\in A} H_\lambda(\sigma) \leq \frac{1}{N}\mathbb{E}\max_{\sigma\in A} H_\lambda(\sigma) + \sqrt{\lambda}\eta_0 \leq \sqrt{\lambda}\big(P(K) - \eta_0\big).$$

This means that, on this event, any σ satisfying

$$\frac{H_\lambda(\sigma)}{N} > \sqrt{\lambda}\big(P(K) - \eta_0\big)$$

must also satisfy $|m(\sigma)| < c = \lambda^{-1/(2K)}(4\eta_0)^{1/K}$. $\square$

5. Overlap gap in the mean field model. In this section, we prove Theorem 3. As mentioned in the Introduction, it will be proved by means of the Guerra–Talagrand (GT) replica symmetry breaking bound at zero temperature, which is obtained from the positive temperature bound [21, 42–44] by an appropriate rescaling procedure. We begin by recalling the analogue of the Parisi formula for the limit of the normalized maximum of the Hamiltonian (1.5),

$$\mathrm{ME}_N := \frac{1}{N}\max_{\sigma\in\{-1,1\}^N} H(\sigma), \tag{5.1}$$

which was proved in [2], as well as the zero-temperature Guerra–Talagrand bound for coupled copies.

5.1. *The Parisi formula and Guerra–Talagrand bound.* Let $\mathcal{U}$ be the set of all nonnegative nondecreasing right-continuous functions γ on the interval $[0, 1)$ such that $\int_0^1 \gamma(s)\,ds < \infty$. We equip the space $\mathcal{U}$ with the $L^1(dx)$ norm. Let

$$\xi(s) = s^K \tag{5.2}$$

and define the zero temperature analogue of the Parisi functional [37, 38] on $\mathcal{U}$ by

$$\mathcal{P}(\gamma) = \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 s\xi''(s)\gamma(s)\,ds, \tag{5.3}$$

where $\Phi_\gamma(s,x)$ for $(s,x)\in[0,1)\times\mathbb{R}$ is the weak[3] solution of the Parisi PDE

$$\partial_s\Phi_\gamma(s,x)=-\frac{\xi''(s)}{2}\big(\partial_{xx}\Phi_\gamma(s,x)+\gamma(s)\big(\partial_x\Phi_\gamma(s,x)\big)^2\big) \tag{5.4}$$

with the boundary condition $\Phi_\gamma(1,x)=|x|$. The existence, uniqueness and regularity properties of the solution $\Phi_\gamma(s,x)$ were studied in [11], Appendix. The Parisi formula for the maximum energy in [2] states

$$\mathrm{ME}:=\lim_{N\to\infty}\mathrm{ME}_N=\inf_{\gamma\in\mathcal{U}}\mathcal{P}(\gamma), \tag{5.5}$$

where the limit of ME_N exists almost surely. Indeed, (5.5) was established in [2] for general mixed p-spin models by showing how the corresponding formulas at positive temperature, first proved for mixed even p-spin models in [43] and for general mixed p-spin models in [33, 34], are transformed in the zero-temperature limit. The minimizer on the right-hand side exists and is unique, as was shown in [11] building upon the ideas from [1]. We will denote this minimizer by γ_P and call $\gamma_P(ds)$ the Parisi measure.

We will now describe the Guerra–Talagrand bound for two coupled systems, which is discussed in Section 3.1 of [11]. For any $(\lambda,\mathbf{x})\in\mathbb{R}\times\mathbb{R}^2$, set

$$g(\lambda,\mathbf{x})=\max(x_1+x_2+\lambda,-x_1-x_2+\lambda,x_1-x_2-\lambda,-x_1+x_2-\lambda). \tag{5.6}$$

Fix $\gamma\in\mathcal{U}$, $q\in[0,1)$ and $\lambda\in\mathbb{R}$. First we define $\Gamma_\gamma(\lambda,s,\mathbf{x})$ for $(s,\mathbf{x})\in[q,1]\times\mathbb{R}^2$ to be the weak solution of the following equation:

$$\partial_s\Gamma_\gamma=-\frac{\xi''(s)}{2}\big(\Delta\Gamma_\gamma+\gamma(s)\langle\nabla\Gamma_\gamma,\nabla\Gamma_\gamma\rangle\big), \tag{5.7}$$

with the boundary condition at $s=1$ given by

$$\Gamma_\gamma(\lambda,1,\mathbf{x})=g(\lambda,\mathbf{x}). \tag{5.8}$$

Here, $\Delta\Gamma_\gamma$ is the Laplacian and $\nabla\Gamma_\gamma$ is the gradient of Γ_γ in $\mathbf{x}$. The existence and uniqueness of the weak solution to this PDE can be established in a similar way as in the one-dimensional case (5.4) without essential changes; see [11] and also [8], Appendix 1. Next, using $\Gamma_\gamma(q,x,x)$ as a boundary condition at $s=q$, we define $\Psi_\gamma(\lambda,s,x)$ for $(s,x)\in[0,q]\times\mathbb{R}$ to be the solution of

$$\partial_s\Psi_\gamma=-\frac{\xi''(s)}{2}\big(\partial_{xx}\Psi_\gamma+\gamma(s)(\partial_x\Psi_\gamma)^2\big) \tag{5.9}$$

[3]Recall from [11, 28] that u is a weak solution to (5.4) if it has an essentially bounded weak derivative $\partial_x u$ and for every $\phi\in C_c^\infty((0,1]\times\mathbb{R})$, $\int_0^1\int_{\mathbb{R}}(-u\partial_t\phi+\frac{\xi''(t)}{2}(u\partial_{xx}\phi+\gamma(s)(\partial_x u)^2))\,dx\,ds+\int_{\mathbb{R}}\phi(1,x)|x|\,dx=0$.

with the boundary condition $\Psi_\gamma(\lambda, q, x) = \Gamma_\gamma(\lambda, q, x, x)$. The existence and regularity properties of Γ_γ and Ψ_γ were studied in [11] and will also be discussed in the proof of Proposition 3 below. Finally, for all $(\lambda, \gamma) \in \mathbb{R} \times \mathcal{U}$, we define

$$\begin{aligned}\mathcal{T}_q(\lambda, \gamma) &= \Psi_\gamma(\lambda, 0, 0) - \lambda q \\ &\quad - \biggl(\int_0^1 s\xi''(s)\gamma(s)\,ds + \int_0^q s\xi''(s)\gamma(s)\,ds\biggr).\end{aligned} \tag{5.10}$$

If for a subset $A \subset [-1, 1]$ we define the maximum coupled energy as

$$\mathrm{MCE}_N(A) = \frac{1}{N} \max_{R_{1,2} \in A} \bigl(H(\sigma^1) + H(\sigma^2)\bigr) \tag{5.11}$$

and let $S_N = \{R_{1,2} : \sigma^1, \sigma^2 \in \{-1, +1\}^N\}$ be the set of feasible overlap values for a given N then, for any $q \in S_N \cap [0, 1]$,

$$\mathbb{E}\mathrm{MCE}_N(\{q\}) \le \mathcal{T}_q(\lambda, \gamma). \tag{5.12}$$

This zero temperature Guerra–Talagrand bound follows from the classical positive temperature bound (see [43] or [42], Section 14.6) by the same rescaling argument as in the setting of one system explained in Lemma 2 in [2]. By symmetry of the model for even K, the same bound holds for $\mathbb{E}\mathrm{MCE}_N(\{-q\})$ as well. Our goal now will be to show that the right-hand side can be made strictly less than twice $\mathcal{P}(\gamma_P)$ defined in (5.5) for all q in some nontrivial interval (a, b), and this will be done by an appropriate choice of γ in (5.12) and slightly perturbing λ around zero. The following subsection prepares needed auxiliary results.

5.2. *Variational representations for Φ_γ and Ψ_γ.* Observe that both Φ_γ and Ψ_γ are special cases of the Hamilton–Jacobi–Bellman equation, induced by a linear problem of diffusion control. In these cases, it is well known that they can be expressed as optimal stochastic control problems by means of dynamic programming [17]. More precisely, denote by $W = \{W(w), \mathcal{G}_w, 0 \le w \le 1\}$ a standard Brownian motion, where the filtration $(\mathcal{G}_w)_{0 \le w \le 1}$ satisfies the usual conditions (see Definition 2.25 in Chapter 1 of [29]). For $0 \le q \le 1$, denote by $D[q]$ the space of all progressively measurable processes u with respect to $(\mathcal{G}_w)_{0 \le w \le q}$ satisfying $\sup_{0 \le w \le q} |u(w)| \le 2$. Endow the space $D[q]$ with the norm

$$\|u\|_q = \biggl(\mathbb{E}\int_0^q u(w)^2\,dw\biggr)^{1/2}. \tag{5.13}$$

The following variational representation holds for Φ_γ.

PROPOSITION 1. *Let $\gamma \in \mathcal{U}$. For any $0 \le s \le q \le 1$ and $x \in \mathbb{R}$,*

$$\begin{aligned}\Phi_\gamma(s, x) = \max_{u \in D[q]} \mathbb{E}\biggl[&\Phi_\gamma\biggl(q, x + \int_s^q \xi''\gamma u\,dw + \int_s^q \sqrt{\xi''}\,dW\biggr) \\ &- \frac{1}{2}\int_s^q \xi''\gamma u^2\,dw\biggr].\end{aligned} \tag{5.14}$$

The maximum is attained on $u_\gamma(w) = \partial_x \Phi_\gamma(w, X_\gamma(w))$, *where* $X_\gamma = (X_\gamma(w))_{s \le w \le q}$ *is the solution to the following SDE*:

$$dX_\gamma = \xi''(w)\gamma(w)\partial_x \Phi_\gamma(w, X_\gamma)\, dw + \sqrt{\xi''(w)}\, dW, \tag{5.15}$$

with the initial condition $X_\gamma(s) = x$.

Similarly, the functional $\Psi_\gamma(\lambda, 0, 0)$ in (5.10) can be written via the following optimal stochastic control problem. Recall that $\Psi_\gamma(\lambda, q, x) = \Gamma_\gamma(\lambda, q, x, x)$.

PROPOSITION 2. *Let* $(\lambda, \gamma) \in \mathbb{R} \times \mathcal{U}$. *For any* $0 \le s \le q \le 1$ *and* $x \in \mathbb{R}$,

$$\Psi_\gamma(\lambda, s, x) = \max_{u \in D[q]} \mathbb{E}\Big[\Psi_\gamma\Big(\lambda, q, x + \int_s^q \xi''\gamma u\, dw + \int_s^q \sqrt{\xi''}\, dW\Big) - \frac{1}{2}\int_s^q \xi''\gamma u^2\, dw\Big]. \tag{5.16}$$

The maximum is attained on $u_\gamma^\lambda(w) = \partial_x \Psi_\gamma(\lambda, w, X_\gamma^\lambda(w))$, *where* $(X_\gamma^\lambda(w))_{s \le w \le q}$ *is the solution of the following SDE*:

$$dX_\gamma^\lambda = \xi''(w)\gamma(w)\partial_x \Psi_\gamma\big(\lambda, w, X_\gamma^\lambda\big)\, dw + \sqrt{\xi''(w)}\, dW, \tag{5.17}$$

with the initial condition $X_\gamma^\lambda(s) = x$.

Alternatively, both Propositions 1 and 2 can also be verified by a direct application of Itô's formula. For a detailed proof, we refer the reader to [10], Theorem 5. Although the argument therein is for a different boundary condition and $\gamma(1-)$ is bounded, the same argument carries through with only minor modifications. Our next result gives a representation of the derivative of $\Psi_\gamma(\lambda, 0, 0)$ in λ when $\lambda = 0$.

PROPOSITION 3. *For any* $0 \le q < 1$ *and any* $\gamma \in \mathcal{U}$, *the function* $\lambda \mapsto \Psi_\gamma(\lambda, 0, 0)$ *is continuously differentiable and its derivative at zero can be written as*

$$\partial_\lambda \Psi_\gamma(0, 0, 0) = \mathbb{E}\partial_x \Phi_\gamma\big(q, X_\gamma^0(q)\big)^2, \tag{5.18}$$

where X_γ^0 *is defined in* (5.17) *with* $\lambda = 0$, $s = 0$ *and* $x = 0$.

The rest of this subsection is devoted to verifying this proposition. Note that a direct computation gives $\partial_\lambda \Gamma_\gamma(0, 1, \mathbf{x}) = \Phi_\gamma(1, x_1) \cdot \partial_{x_2} \Phi_\gamma(1, x_2)$ for all $x_1, x_2 \ne 0$. From this, the validity of Proposition 3 follows immediately from the Cole–Hopf transformation so long as $\gamma \in \mathcal{U}$ is a step function with finitely many jumps. To handle the general situation, we need the following three key lemmas, which relate the functionals in the Guerra–Talagrand bound and the Parisi formula as well as describe their regularity properties. They are taken directly from [10] and [11]. The first one is in Lemma 6 in [11].

LEMMA 5.1. *Fix $0 \le q < 1$ and $\gamma \in \mathcal{U}$. For any $w \in [q, 1)$ and $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$,*

$$\Gamma_\gamma(0, w, \mathbf{x}) = \Phi_\gamma(w, x_1) + \Phi_\gamma(w, x_2) \tag{5.19}$$

and $\Gamma_\gamma(\lambda, w, \mathbf{x})$ is differentiable in λ in the classical sense with

$$\partial_\lambda \Gamma_\gamma(0, w, \mathbf{x}) = \partial_x \Phi_\gamma(w, x_1)\partial_x \Phi_\gamma(w, x_2). \tag{5.20}$$

The second one can be found in Lemma 2 in [10].

LEMMA 5.2. *Let $\mathcal{D}$ be a metric space and I be an interval with right open edge. Let f be a real-valued function on $\mathcal{D} \times I$ and $f_0(\lambda) = \sup_{u \in \mathcal{D}} f(u, \lambda)$ for $\lambda \in I$. Suppose that there exists a $\mathcal{D}$-valued continuous function $u(\lambda)$ on I such that $f_0(\lambda) = f(u(\lambda), \lambda)$ and $\partial_\lambda f$ is continuous on $\mathcal{D} \times I$. Then $f_0(\lambda)$ is right-differentiable with the derivative given by $\partial_\lambda f(u(\lambda), \lambda)$ for all $\lambda \in I$.*

We will also need the following Lemma 10 in [11].

LEMMA 5.3. *Let $0 < r_0 < r_1 < r_2 < \infty$. Suppose that $\kappa_1, \kappa_2 \in L^\infty([r_0, r_2] \times \mathbb{R})$ and $g \in L^\infty(\mathbb{R})$ with $\|\kappa_i\|_\infty \le C_i$ for $i = 1, 2$ and $\|g\|_\infty \le C_0$. Assume that u is the classical solution to*

$$\partial_r u(r, x) = \partial_{xx} u(r, x) + \kappa_1(r, x)\partial_x u(r, x) + k_2(r, x) \qquad \forall (r, x) \in (r_0, r_2] \times \mathbb{R}$$

with the initial condition $u(r_0, x) = g(x)$. Then there exists a nonnegative continuous function F on $[0, \infty)^3$ depending only on r_0, r_2 such that

$$\sup_{(r,x) \in [r_1, r_2] \times \mathbb{R}} \left|\partial_x u(r, x)\right| \le F(C_0, C_1, C_2).$$

The statement of this result in [11] also assumed the continuity of $r \mapsto \|\partial_x u(r, \cdot)\|_\infty$. To avoid this, we note that this assumption was used in [11] only in the application of Gronwall's inequality, which also holds without it, so this result in [11] in fact did not need this assumption.

PROOF OF PROPOSITION 3. First we establish the differentiability of $\Psi_\gamma(\lambda, 0, 0)$. Again, recall that $\Psi_\gamma(\lambda, q, x) = \Gamma_\gamma(\lambda, q, x, x)$. If we denote

$$F_\gamma^q(\lambda, u) = \mathbb{E}\left[\Psi_\gamma\left(\lambda, q, \int_0^q \xi'' \gamma u \, dw + \int_0^q \sqrt{\xi''} dW\right) - \frac{1}{2}\int_0^q \xi'' u \, dw\right],$$

for $u \in D[q]$ then, by Proposition 2,

$$\Psi_\gamma(\lambda, 0, 0) = \max_{u \in D[q]} F_\gamma^q(\lambda, u).$$

To compute the derivative of this in λ, we would like to apply Lemma 5.2 with $I = \mathbb{R}$, the metric space $\mathcal{D} = (D[q], \|\cdot\|_q)$, and $u(\lambda) = u_\gamma^\lambda$ from Proposition 2.

First of all, the proof of Lemma 6 in [11] shows that $\Psi_\gamma(\lambda,q,x)=\Gamma_\gamma(\lambda,q,x,x)$ is differentiable in λ, the derivative $\partial_\lambda\Psi_\gamma(\lambda,q,x)$ is continuous in (λ,x), and $|\partial_\lambda\Psi_\gamma(\lambda,q,x)|\le 1$. In particular, this implies that we can interchange the derivative and integral and write

$$\partial_\lambda F_\gamma^q(\lambda,u)=\mathbb{E}\biggl[\partial_\lambda\Psi_\gamma\Bigl(\lambda,q,\int_0^q\xi''\gamma u\,dw+\int_0^q\sqrt{\xi''}\,dW\Bigr)\biggr]. \tag{5.21}$$

For $u(\lambda)=u_\gamma^\lambda$ from Proposition 2 with $x=0$ and $s=0$,

$$\int_0^q\xi''\gamma u_\gamma^\lambda\,dw+\int_0^q\sqrt{\xi''}\,dW=X_\gamma^\lambda(q)$$

and, therefore, if we can show that $\partial_\lambda F_\gamma^q(\lambda,u)$ is continuous in (λ,u) and u_γ^λ is continuous in λ on $\mathcal{D}$ then Lemma 5.2 will imply that the right derivative

$$\frac{d}{d\lambda}\Psi_\gamma(\lambda,0,0)=\partial_\lambda F_\gamma^q\bigl(\lambda,u_\gamma^\lambda\bigr)=\mathbb{E}\bigl[\partial_\lambda\Psi_\gamma\bigl(\lambda,q,X_\gamma^\lambda(q)\bigr)\bigr]. \tag{5.22}$$

Since $\partial_\lambda\Psi_\gamma$ is continuous in (λ,x) and $|\partial_\lambda\Psi_\gamma|\le 1$, (5.21) implies that $\partial_\lambda F_\gamma^q$ is continuous in (λ,u).

It remains to show that u_γ^λ is continuous in λ on $\mathcal{D}$. First we note that $\Gamma_\gamma(\lambda,q,x,x)$ admits a similar expressions as Propositions 1 and 2, which can be seen from Theorem 7 in [11] with $(r,s)=(q,1)$ and $t=1$. This representation makes it clear how properties of the boundary condition $g(\lambda,\mathbf{x})$ in (5.8) are propagated to $\Psi_\gamma(\lambda,q,x)$. For example, the fact that $g(\lambda,\mathbf{x})$ is 1-Lipschitz in λ and 2-Lipschitz in $\mathbf{x}$ implies that $|\partial_\lambda\Psi_\gamma(\lambda,q,x)|\le 1$ and $|\partial_x\Psi_\gamma(\lambda,q,x)|\le 2$. Using (5.16), we get $|\partial_\lambda\Psi_\gamma(\lambda,s,x)|\le 1$ and $|\partial_x\Psi_\gamma(\lambda,s,x)|\le 2$ for all $s\in[0,q]$. From this, it can be shown that $\partial_x\Psi_\gamma(\lambda,s,x)$ is uniformly Lipschitz in (λ,x) over all $s\in[0,q]$, as follows.

If $\gamma(ds)$ is an atomic measure with finitely many jumps then the solution Ψ_γ of (5.9) can be explicitly found via the Cole–Hopf transformation. Moreover, $A:=\partial_x\Psi_\gamma$ or $A:=\partial_\lambda\Psi_\gamma$ satisfies

$$\partial_t\Delta=-\frac{\xi''}{2}\bigl(\partial_{xx}\Delta+2\gamma(\partial_x\Psi_\gamma)(\partial_x\Delta)\bigr)$$

on $\mathbb{R}\times[0,q]\times\mathbb{R}$ with the boundary condition $A(\lambda,q,x)=\partial_x\Psi_\gamma(\lambda,q,x)$ or $\partial_\lambda\Psi_\gamma(\lambda,q,x)$. Since $|A(\lambda,q,x)|\le 2$ and $|\partial_x\Psi_\gamma(\lambda,s,x)|\le 2$ for $s\in[0,q]$, one may apply Lemma 5.3 to obtain

$$\sup_{(\lambda,s,x)\in\mathbb{R}\times[0,q]\times\mathbb{R}}\bigl|\partial_x A(\lambda,s,x)\bigr|\le F\bigl(\gamma(q)\bigr), \tag{5.23}$$

where F is some nonnegative continuous function on $[0,\infty)$. Therefore, $\partial_x\Psi_\gamma(\lambda,s,x)$ is $F(\gamma(q))$-Lipschitz in λ and x uniformly over $s\in[0,q]$, for atomic $\gamma(ds)$. For arbitrary γ on $[0,q]$, we define the solution $\Psi_\gamma(\lambda,s,x)$ of (5.9) with the boundary condition $\Gamma_\gamma(\lambda,q,x,x)$ via approximation of γ by atomic γ_n

and ensuring, using the Arzela–Ascoli theorem and Cantor's diagonalization process, that $\Psi_{\gamma_n}(\lambda, s, x)$ and all its partial derivatives in x and partial derivative in λ converge to some $\Psi_\gamma(\lambda, s, x)$ and its partial derivatives. The limiting $\Psi_\gamma(\lambda, s, x)$ is the solution of the equation (5.9) with the boundary condition $\Gamma_\gamma(\lambda, q, x, x)$ and (5.23) still holds, that is, $\partial_x \Psi_\gamma(\lambda, s, x)$ is $F(\gamma(q))$-Lipschitz in λ and x uniformly over $s \in [0, q]$. (The uniqueness of solutions follows from the same argument as in [28].)

Finally, equation (5.17) with $x = 0$ and $s = 0$ and (5.23) imply that, for $0 \le w \le q$,

$$\left|X_\gamma^\lambda(w) - X_\gamma^{\lambda'}(w)\right| \le \xi''(q)\gamma(q)F\big(\gamma(q)\big)\left(|\lambda - \lambda'|w + \int_0^w \left|X_\gamma^\lambda(t) - X_\gamma^{\lambda'}(t)\right| dt\right).$$

An application of Gronwall's inequality shows that $\lambda \mapsto X_\gamma^\lambda$ is continuous with respect to the sup-norm, and (5.23) implies that $\lambda \mapsto u_\gamma^\lambda(w) = \partial_x \Psi_\gamma(\lambda, w, X_\gamma^\lambda(w))$ is also continuous with respect to the sup-norm and, therefore, the norm $\|\cdot\|_q$ defined in (5.15). As a result, $\lambda \mapsto (\lambda, u_\gamma^\lambda)$ is continuous and (5.22) holds. Note that this right derivative is also continuous in λ. It is well known (see, e.g., [9]) that a function with continuous right derivative on an open interval is continuously differentiable on this interval. Thus, $\Psi_\gamma(\lambda, 0, 0)$ is continuously differentiable. Finally, the verification of (5.18) follows directly from (5.22) and (5.20). □

5.3. *Proof of Theorem 3.* We now turn to the main proof of Theorem 3 by using the GT bound (5.12). First of all, from the optimality of the Parisi measure γ_P, it is known that the points in the support of the Parisi measure must satisfy a system of consistency equations (see [11], Proposition 3).

PROPOSITION 4. *For any $s \in \operatorname{supp} \gamma_P(ds)$,*

$$\mathbb{E}\big(\partial_x \Phi_{\gamma_P}(s, X_{\gamma_P}(s))\big)^2 = s, \tag{5.24}$$

$$\xi''(s)\mathbb{E}\big(\partial_{xx} \Phi_{\gamma_P}(s, X_{\gamma_P}(s))\big)^2 \le 1, \tag{5.25}$$

where $X_{\gamma_P} = (X_{\gamma_P}(w))_{0 \le w < 1}$ is defined as in Proposition 1 with $q = 1$, $s = 0$ and $x = 0$.

A key consequence of these equations is the following lemma where the condition $K \ge 4$ is used.

LEMMA 5.4. *For even $K \ge 4$, there exists $c \in (0, 1)$ such that*

$$\mathbb{E}\big(\partial_x \Phi_{\gamma_P}(s, X_{\gamma_P}(s))\big)^2 < s$$

for all $s \in (0, c)$.

PROOF. Note that $\xi(s) = s^K$. By Proposition 2(ii) in [11], for any $0 < q < 1$, there exists a nonnegative continuous function F on $[0, \infty)$ depending only on ξ and q such that

$$\sup_{(s,x)\in[0,q]\times\mathbb{R}} \bigl|\partial_{xx}\Phi_{\gamma_P}(s,x)\bigr| \le F\bigl(\gamma_P(q)\bigr).$$

Since $\xi''(0) = 0$ for $K \ge 4$, this and (5.25) imply that there exists some $c \in (0,1)$ such that

$$\xi''(s)\mathbb{E}\bigl(\partial_{xx}\Phi_{\gamma_P}\bigl(s, X_{\gamma_P}(s)\bigr)\bigr)^2 < 1 \tag{5.26}$$

for all $s \in (0, c)$. It remains to appeal to the following well-known relationship between the left-hand sides of the equations (5.24) and (5.24). By [11], Lemma 3,

$$\begin{aligned}&\partial_x\Phi_{\gamma_P}\bigl(s, X_\gamma(s)\bigr) - \partial_x\Phi_{\gamma_P}\bigl(s', X_\gamma(s')\bigr)\\ &\quad= \int_{s'}^{s} \xi''(w)^{1/2}\partial_{xx}\Phi_{\gamma_P}\bigl(w, X_{\gamma_P}(w)\bigr)\,dW(w)\end{aligned}$$

for all $0 < s' < s < 1$, and application of Itô's formula yields

$$\begin{aligned}&\mathbb{E}\bigl(\partial_x\Phi_{\gamma_P}\bigl(s, X_{\gamma_P}(s)\bigr)\bigr)^2 - \mathbb{E}\bigl(\partial_x\Phi_{\gamma_P}\bigl(s', X_{\gamma_P}(s')\bigr)\bigr)^2\\ &\quad= \int_{s'}^{s} \xi''(w)\mathbb{E}\bigl(\partial_{xx}\Phi_{\gamma_P}\bigl(w, X_{\gamma_P}(w)\bigr)\bigr)^2\,dw.\end{aligned}$$

In particular, letting $s' = 0$ and using (5.26),

$$\mathbb{E}\bigl(\partial_x\Phi_{\gamma_P}\bigl(s, X_{\gamma_P}(s)\bigr)\bigr)^2 < s$$

for any $s \in (0, c)$. This completes our proof. □

Let c be the constant obtained in Lemma 5.4. For any fixed $q \in (0, c)$, we define

$$\gamma_q(s) = \frac{\gamma_P}{2}1_{[0,q)} + \gamma_P 1_{[q,1)} \in \mathcal{U},$$

where 1_A stands for the indicator function on the set A. The definition of the boundary condition in (5.9), $\Psi_\gamma(\lambda, q, x) = \Gamma_\gamma(\lambda, q, x, x)$, together with (5.19) yields

$$\Psi_{\gamma_q}(0, q, x) = \Gamma_{\gamma_q}(0, q, x, x) = 2\Phi_{\gamma_P}(q, x).$$

Since $\gamma_q = \gamma_P/2$ on the interval $[0, q)$, comparing the equations (5.4) and (5.9) we see that

$$\Psi_{\gamma_q}(0, s, x) = 2\Phi_{\gamma_P}(s, x) \qquad \text{for } s \in [0, q]. \tag{5.27}$$

Moreover,

$$\int_0^1 s\xi''(s)\gamma_q(s)\,ds + \int_0^q s\xi''(s)\gamma_q(s)\,ds = \int_0^1 s\xi''(s)\gamma_P(s)\,ds$$

and, as a result,

$$\mathcal{T}_q(0,\gamma_q) = 2\mathcal{P}(\gamma_P) = 2\mathrm{ME}. \tag{5.28}$$

Next, recall X_{γ_q} from (5.15) and $X^0_{\gamma_q}$ from (5.17) both with $x = 0$ and $s = 0$. Since on the interval $[0, q]$ we have $\gamma_q = \gamma_P/2$ and $\partial_x \Psi_{\gamma_q}(0, s, x) = 2\partial_x \Phi_{\gamma_P}(s, x)$, (5.17) becomes

$$dX^0_{\gamma_q} = \xi'' \gamma_P \partial_x \Phi_{\gamma_q}\big(s, X^0_{\gamma_q}\big)\, ds + \sqrt{\xi''}\, dW.$$

In other words, X_{γ_q} and $X^0_{\gamma_q}$ are defined by the same equation on $[0, q]$ and, by the uniqueness of the strong SDE solution, we see that $X^0_{\gamma_q} = X_{\gamma_P}$. Consequently, (5.18) implies

$$\partial_\lambda \Psi_{\gamma_q}(0,0,0) = \mathbb{E}\partial_x \Phi_{\gamma_P}\big(q, X_{\gamma_P}(q)\big)^2$$

and, by Lemma 5.4,

$$\partial_\lambda \mathcal{T}_q(0,\gamma_q) = \partial_\lambda \Psi_{\gamma_q}(0,0,0) - q = \mathbb{E}\partial_x \Phi_{\gamma_P}\big(q, X_{\gamma_P}(q)\big)^2 - q < 0.$$

This assures that, for all $q \in (0, c)$,

$$\inf_{(\lambda,\gamma)\in\mathbb{R}\times\mathcal{U}} \mathcal{T}_q(\lambda,\gamma) < 2ME.$$

Since $\mathcal{T}_q(\lambda,\gamma)$ is continuous in q, the infimum on the left-hand side is an upper semicontinuous function of q and, therefore, there exist $0 < a < b < c$ and $\eta > 0$ such that

$$\inf_{(\lambda,\gamma)\in\mathbb{R}\times\mathcal{U}} \mathcal{T}_q(\lambda,\gamma) < 2ME - 2\eta$$

for all $q \in [a, b]$. Consequently, from (5.12) and the usual Gaussian concentration of measure,

$$\mathbb{E}\mathrm{MCE}_N\big([a,b]\big) \le 2\mathbb{E}\mathrm{ME}_N - \eta$$

for large enough N. Note that, since the Hamiltonian $H(\sigma)$ is symmetric for even K, that is, $H(\sigma) = H(-\sigma)$, we see that this also implies that

$$\mathbb{E}\mathrm{MCE}_N\big([-b,-a]\cup[a,b]\big) \le 2\mathbb{E}\mathrm{ME}_N - \eta.$$

This completes our proof. □

Acknowledgments. The authors thank a referee for a careful reading and for giving several suggestions regarding the presentation of the paper. David Gamarnik and Mustazee Rahman also thank organizers of the AIM workshop "Phase transitions in randomized computational problems" where some of this work was conducted.

REFERENCES

[1] AUFFINGER, A. and CHEN, W.-K. (2015). The Parisi formula has a unique minimizer. *Comm. Math. Phys.* **335** 1429–1444. MR3320318

[2] AUFFINGER, A. and CHEN, W.-K. (2017). Parisi formula for the ground state energy in the mixed p-spin model. *Ann. Probab.* **45** 4617–4631. MR3737919

[3] AUFFINGER, A., CHEN, W.-K. and ZENG, Q. (2017). The SK model is Full-step Replica Symmetry Breaking at zero temperature. Preprint. Available at arXiv:1703.06872.

[4] BACKHAUSZ, A. and SZEGEDY, B. (2016). On the almost eigenvectors of random regular graphs. Preprint. Available at arXiv:1607.04785.

[5] BACKHAUSZ, A. and SZEGEDY, B. (2018). On large girth regular graphs and random processes on trees. *Random Structures Algorithms*. To appear. Available at arXiv:1406.4420.

[6] BACKHAUSZ, Á. and VIRÁG, B. (2017). Spectral measures of factor of i.i.d. processes on vertex-transitive graphs. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2260–2278. MR3729654

[7] BAYATI, M., GAMARNIK, D. and TETALI, P. (2013). Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.* **41** 4080–4115. MR3161470

[8] BEN AROUS, G. and JAGANNATH, A. (2018). Spectral gap estimates in mean field spin glasses. *Comm. Math. Phys.* **361** 1–52. MR3825934

[9] BRUCKNER, A. (1994). *Differentiation of Real Functions*, 2nd ed. *CRM Monograph Series* **5**. Amer. Math. Soc., Providence, RI. MR1274044

[10] CHEN, W.-K. (2017). Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *Ann. Probab.* **45** 3929–3966. MR3729619

[11] CHEN, W.-K., HANDSCHY, M. and LERMAN, G. (2018). On the energy landscape of the mixed even p-spin model. *Probab. Theory Related Fields* **171** 53–95. MR3800830

[12] CHEN, W.-K. and PANCHENKO, D. (2018). Disorder chaos in some diluted spin glass models. *Ann. Appl. Probab.* **28** 1356–1378. MR3809466

[13] DE SANCTIS, L. (2004). Random multi-overlap structures and cavity fields in diluted spin glasses. *J. Stat. Phys.* **117** 785–799. MR2107895

[14] DEMBO, A., MONTANARI, A. and SEN, S. (2017). Extremal cuts of sparse random graphs. *Ann. Probab.* **45** 1190–1217. MR3630296

[15] ELEK, G. and LIPPNER, G. (2010). Borel oracles. An analytical approach to constant-time algorithms. *Proc. Amer. Math. Soc.* **138** 2939–2947. MR2644905

[16] FAN, Z. and MONTANARI, A. (2017). How well do local algorithms solve semidefinite programs? In *STOC'*17*—Proceedings of the* 49*th Annual ACM SIGACT Symposium on Theory of Computing* 604–614. ACM, New York. MR3678214

[17] FLEMING, W. H. and SONER, H. M. (2006). *Controlled Markov Processes and Viscosity Solutions*, 2nd ed. *Stochastic Modelling and Applied Probability* **25**. Springer, New York. MR2179357

[18] FRANZ, S. and LEONE, M. (2003). Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.* **111** 535–564. MR1972121

[19] GAMARNIK, D. and SUDAN, M. (2017). Limits of local algorithms over sparse random graphs. *Ann. Probab.* **45** 2353–2376. MR3693964

[20] GAMARNIK, D. and SUDAN, M. (2017). Performance of sequential local algorithms for the random NAE-K-SAT problem. *SIAM J. Comput.* **46** 590–619. MR3620150

[21] GUERRA, F. (2003). Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233** 1–12. MR1957729

[22] GUERRA, F. and TONINELLI, F. L. (2002). The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.* **230** 71–79. MR1930572

[23] GUERRA, F. and TONINELLI, F. L. (2004). The high temperature region of the Viana–Bray diluted spin glass model. *J. Stat. Phys.* **115** 531–555. MR2070106

[24] HARANGI, V. and VIRÁG, B. (2015). Independence ratio and random eigenvectors in transitive graphs. *Ann. Probab.* **43** 2810–2840. MR3395475

[25] HATAMI, H., LOVÁSZ, L. and SZEGEDY, B. (2014). Limits of locally-globally convergent graph sequences. *Geom. Funct. Anal.* **24** 269–296. MR3177383

[26] HOPPEN, C. and WORMALD, N. (2013). Local algorithms, regular graphs of large girth, and random regular graphs. Preprint. Available at arXiv:1308.0266.

[27] JAGANNATH, A., KO, J. and SEN, S. (2018). Max κ-cut and the inhomogeneous Potts spin glass. *Ann. Appl. Probab.* **28** 1536–1572. MR3809471

[28] JAGANNATH, A. and TOBASCO, I. (2016). A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.* **144** 3135–3150. MR3487243

[29] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940

[30] LYONS, R. (2017). Factors of IID on trees. *Combin. Probab. Comput.* **26** 285–300. MR3603969

[31] LYONS, R. and NAZAROV, F. (2011). Perfect matchings as IID factors on non-amenable groups. *European J. Combin.* **32** 1115–1125. MR2825538

[32] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick Model*. Springer, New York. MR3052333

[33] PANCHENKO, D. (2013). The Parisi ultrametricity conjecture. *Ann. of Math.* (2) **177** 383–393. MR2999044

[34] PANCHENKO, D. (2014). The Parisi formula for mixed p-spin models. *Ann. Probab.* **42** 946–958. MR3189062

[35] PANCHENKO, D. (2018). On the K-sat model with large number of clauses. *Random Structures Algorithms* **52** 536–542. MR3783209

[36] PANCHENKO, D. and TALAGRAND, M. (2004). Bounds for diluted mean-fields spin glass models. *Probab. Theory Related Fields* **130** 319–336. MR2095932

[37] PARISI, G. (1980). A sequence of approximate solutions to the S-K model for spin glasses. *J. Phys. A* **13** L115–L121.

[38] PARISI, G. (1979). Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.* **43** 1754–1756.

[39] RAHMAN, M. (2016). Factor of IID percolation on trees. *SIAM J. Discrete Math.* **30** 2217–2242. MR3578023

[40] RAHMAN, M. and VIRÁG, B. (2017). Local algorithms for independent sets are half-optimal. *Ann. Probab.* **45** 1543–1577. MR3650409

[41] SEN, S. (2016). Optimization on sparse random hypergraphs and spin glasses. *Random Structures Algorithms*. To appear. Available at arXiv:1606.02365.

[42] TALAGRAND, M. (2011). *Mean-Field Models for Spin Glasses. Ergebnisse der Mathematik und ihrer Grenzgebiete*. 3. *Folge A Series of Modern Surveys in Mathematics* **54**. Springer, Berlin.

[43] TALAGRAND, M. (2006). The Parisi formula. *Ann. of Math.* (2) **163** 221–263. MR2195134

[44] TALAGRAND, M. (2007). Mean field models for spin glasses: Some obnoxious problems. In *Spin Glasses. Lecture Notes in Math.* **1900** 63–80. Springer, Berlin. MR2309598

W.-K. CHEN
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA
USA
E-MAIL: wkchen@umn.edu

D. GAMARNIK
SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
USA
E-MAIL: gamarnik@mit.edu

D. PANCHENKO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
TORONTO, ONTARIO
CANADA
E-MAIL: panchenk@math.toronto.edu

M. RAHMAN
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
USA
E-MAIL: mustazee@mit.edu