

# PHASE TRANSITION IN THE SPIKED RANDOM TENSOR WITH RADEMACHER PRIOR<sup>1</sup>

BY WEI-KUO CHEN

*University of Minnesota*

We consider the problem of detecting a deformation from a symmetric Gaussian random  $p$ -tensor ( $p \geq 3$ ) with a rank-one spike sampled from the Rademacher prior. Recently, in Lesieur et al. (Barbier, Krzakala, Macris, Miolane and Zdeborová (2017)), it was proved that there exists a critical threshold  $\beta_p$  so that when the signal-to-noise ratio exceeds  $\beta_p$ , one can distinguish the spiked and unspiked tensors and weakly recover the prior via the minimal mean-square-error method. On the other side, Perry, Wein and Bandeira (Perry, Wein and Bandeira (2017)) proved that there exists a  $\beta'_p < \beta_p$  such that any statistical hypothesis test cannot distinguish these two tensors, in the sense that their total variation distance asymptotically vanishes, when the signal-to-noise ratio is less than  $\beta'_p$ . In this work, we show that  $\beta_p$  is indeed the critical threshold that strictly separates the distinguishability and indistinguishability between the two tensors under the total variation distance. Our approach is based on a subtle analysis of the high temperature behavior of the pure  $p$ -spin model with Ising spin, arising initially from the field of spin glasses. In particular, we identify the signal-to-noise criticality  $\beta_p$  as the critical temperature, distinguishing the high and low temperature behavior, of the Ising pure  $p$ -spin mean-field spin glass model.

**1. Introduction.** The problem of detecting a deformation in a symmetric Gaussian random tensor is concerned about whether there exists a statistical hypothesis test that can reliably distinguish the deformation from the noise. In the matrix case, if  $W$  is a Gaussian Wigner ensemble and  $u$  is a unit vector, the goal is to distinguish the unspiked matrix  $W$  and the spiked matrix  $W + \beta uu^T$  for a given signal-to-noise ratio  $\beta$ . It is well known that in this case the top eigenvalue of the spiked matrix exhibits the so-called BBP (Baik, Ben Arous and P ech e) transition [3, 13, 15, 41]. Namely, the top eigenvalue successfully detects the signal if the strength of  $\beta$  exceeds the critical threshold 1, while it fails to provide indicative information if  $\beta < 1$ . It was further proved in [32, 35, 43] that in the case of spherical, central Gaussian and Rademacher priors, every statistical hypothesis test cannot reliably distinguish the spiked and unspiked matrices.

---

Received December 2017; revised August 2018.

<sup>1</sup>Supported in part by NSF DMS-1642207, NSF DMS-1752184 and Hong Kong Research Grants Council GRF-14302515.

*MSC2010 subject classifications.* Primary 93E10; secondary 60K35, 82B44.

*Key words and phrases.* BBP transition, signal detection, Parisi formula, replica symmetry breaking, spin glass, spiked tensor.

In recent years, the above phenomenon was also studied in the spiked symmetric Gaussian random  $p$ -tensor model of the form,  $T = W + \beta u^{\otimes p}$ . Earlier results were obtained by Montanari and Richard [33] and Montanari, Reichman and Zeitouni [32] in the setting of spherical prior, where they showed that there exist  $\beta_p^-$  and  $\beta_p^+$  with  $\beta_p^- < \beta_p^+$  such that if the signal-to-noise ratio exceeds  $\beta_p^+$ , it is possible to distinguish the spiked and unspiked tensors and weakly recover the signal, but these are impossible if the signal-to-noise ratio is less than  $\beta_p^-$ . More recently, Lesieur et al. [29] studied this detection problem for more general priors by means of the minimal mean-square-error (MMSE), that is,

$$\text{MMSE}_N(\beta) = \inf_{\mathcal{E}} \sum_{1 \leq i_1, \dots, i_p \leq N} \mathbb{E}(\beta u_{i_1} \cdots u_{i_p} - \mathcal{E}_{i_1, \dots, i_p}(T))^2,$$

where the infimum is taken over all bounded measurable functions  $\mathcal{E} = (\mathcal{E}_{i_1, \dots, i_p})$  defined on the space of symmetric tensors. Here, evidently the minimizer to this problem is attained by the MMSE estimator

$$\hat{\mathcal{E}}_{i_1, \dots, i_p}(T) := \mathbb{E}[u_{i_1} \cdots u_{i_p} | T].$$

In [29], it was proved that there exists a critical threshold  $\beta_p^{\text{MMSE}}$  (depending on the prior) so that when  $\beta < \beta_p^{\text{MMSE}}$ , the MMSE estimator fails to distinguish the two tensors and in fact it is no better than a random guess under the mean square error, while for  $\beta > \beta_p^{\text{MMSE}}$ , detection is possible since now MMSE estimator has the best performance among all possible choices. On the other side, Perry, Wein and Bandeira [42] studied the detection problem with spherical, Rademacher, and sparse Rademacher priors. In the case of spherical prior, they provided an improvement on the bounds in [32, 33]. Moreover, in the latter two cases, their results showed that there exists a  $\beta'_p < \beta_p^{\text{MMSE}}$  such that for any  $\beta < \beta'_p$ , it is impossible to distinguish the two tensors in the sense that the total variation distance between the spiked and unspiked tensors asymptotically vanishes. As a consequence, every statistical hypothesis test fails to distinguish the two tensors. The paper [42] then left with a conjecture that indistinguishability between the two tensors should be valid up to the critical threshold  $\beta_p^{\text{MMSE}}$  for both the Rademacher and sparse Rademacher priors.

The aim of this paper is to study the symmetric Gaussian random  $p$ -tensor ( $p \geq 3$ ) with Rademacher prior as in the setting of [42]. We show that there exists a critical value  $\beta_p$  that strictly separates the distinguishability and indistinguishability between the spiked and unspiked tensors under the total variation distance. More precisely, it is established that when the signal-to-noise ratio is less than the critical value  $\beta_p$ , the total variation distance between the spiked and unspiked tensors converges to zero. This establishes the aforementioned prediction in [42]. In particular, we identify the critical value  $\beta_p$  as the critical temperature, distinguishing the high and low temperature behavior, of the Ising pure  $p$ -spin mean-field spin

glass model. This constant also agrees with the critical threshold  $\beta_p^{\text{MMSE}}$  suggested by the MMSE method when applying to the Rademacher prior.

Our approach is based on the methodologies originated from the study of mean-field spin glasses, especially those for the Sherrington–Kirkpatrick model and the mixed  $p$ -spin models; see [37, 46, 47]. Roughly speaking, spin glass models are disordered spin systems initially invented by theoretical physicists in order to explain the strange magnetic behavior of certain alloys, such as CuMn. Mathematically, they are usually formulated as stochastic processes with high complexity and present several crucial features, for example, quenched disorder and frustration, that are commonly shared in many real world problems, involving randomized combinatorial optimization. Over the past decades, the study of spin glasses has received a lot of attention in both physics and mathematics communities; see [30] for a physics overview and [37, 46, 47] for mathematical development.

One way to investigate the detection problem in the symmetric Gaussian random tensor is through the total variation distance between the spiked and unspiked tensors. While in the detection problem  $\beta$  represents the signal-to-noise ratio, we regard  $\beta$  as a (inverse) temperature parameter in the pure  $p$ -spin model. Notably, under this setting, the ratio of the densities between the two tensors can be computed as the partition function of the pure  $p$ -spin model with temperature  $\beta$ . In [32, 42], the authors controlled the total variation distance by the second moment of the partition function. Different than their consideration, we relate this distance to the free energy of the pure  $p$ -spin model with Ising spin; see Lemma 3.2. This relation allows us to show that the critical threshold  $\beta_p$  can be determined by the critical temperature of the pure  $p$ -spin model. In bounding the total variation distance, the most critical ingredient is played by a sharp upper bound concerning the fluctuation of the free energy up to the critical temperature for all  $p \geq 3$ . In the case  $p = 2$ , the pure  $p$ -spin model is famously known as the Sherrington–Kirkpatrick model and its free energy was shown to possess a Gaussian central limit theorem in the weak limit up to the critical temperature  $\beta_p = 1$  by Aizenman, Lebowitz and Ruelle [1]. As for even  $p \geq 4$ , Bovier, Kurkova and Löwe [14] showed that the same result also holds (with different scaling than that in the Sherrington–Kirkpatrick model), but not up to the critical temperature. Our main contribution is that we obtain a sharp upper bound for the fluctuation of the free energy, which is comparable to the one in [14] and more importantly it is valid up to the critical temperature for all  $p \geq 3$  including odd  $p$ . This allows us to extract a sharp upper bound for the total variation distance and deduce the desired result.

Besides the consideration of the detection problem, we also present some new results and arguments for the pure  $p$ -spin models that are of independent interest in spin glasses. First, we show that if the temperature is below the critical value  $\beta_p$ , the model presents the high temperature or *replica symmetric* solution in the sense that any two independently sampled spin configurations from the Gibbs measure are essentially orthogonal to each other by providing exponential tail probability and moment controls. While these results can also be established at *very* high

temperature by some well-known techniques in spin glasses, such as the cavity method, the second moment method, and Latala's argument (see [46], Chapter 1), it is relatively a more challenging task to obtain the same behavior throughout the entire high temperature regime. We show that this is achievable in the pure  $p$ -spin model (see Theorems 4.2 and 4.3) and indeed, our method can also be applied to more general situations, the mixed  $p$ -spin models (see Remark 4.1). Next, in terms of technicality, our argument for the above result is based on the Guerra–Talagrand replica symmetry breaking bound for the coupled free energy for two systems. This bound has been playing a critical role in the study of the mixed *even*  $p$ -spin models; see Talagrand [47]. Its validity for the model involving odd  $p$  mixture is however generally unknown as it is unclear whether the error term along Guerra's replica symmetry breaking interpolation possesses a nonnegative sign or not. To tackle this obstacle, we adopt the synchronization property, introduced by Panchenko [39, 40], that the overlap matrix is asymptotically symmetric and positive semidefinite under the Gibbs average, which was established heavily relying on the fact that the Ghirlanda–Guerra identities imply ultrametricity of the overlaps [36]. This allows us to show that the error term creates a nonnegative sign and ultimately leads to the validity of the Guerra–Talagrand bound in the pure odd  $p$ -spin model if one restricts the functional order parameters to be of *one-step replica symmetry breaking*. Whether this bound is valid for more general functional order parameters remains open.

For other related works on the detection problem of spiked matrices and tensors, we invite the readers to check a variety of low rank matrix estimation problems, including explicit characterizations of mutual information in [5, 26–28, 31] and the performance of the approximate message passing (AMP) for MMSE method and compressed sensing in [5, 11, 21, 22, 25]. More recently, the fluctuation of the likelihood ratio in the spiked Wigner model was studied in [23]. In the case of spiked tensors, phase transitions of the mutual information and the MMSE estimator were recently studied for any  $p$  and given prior; see [7, 29] for symmetric case and [9] for non-symmetric case. The performance of the AMP in the spiked tensor was also investigated in [29]; see also [4, 6, 8, 19, 20, 34, 44]. For the study of Gaussian random  $p$ -tensor in terms of complexity, see [12].

This paper is organized as follows. In Section 2, we state our main results on the detection problem. In Section 3, their proofs are presented and are essentially self-contained for those who wish to learn only the roles of spin glass results in the detection problem. The high temperature results on the pure  $p$ -spin model are all gathered in Section 4. Their proofs are deferred to the Supplementary Material [18] with great details.

**2. Main results.** We begin by setting some standard notation. Let  $p \in \mathbb{N}$ . For any  $N \geq 1$ , denote by  $\Omega_N := \otimes_{j=1}^p \mathbb{R}^N$  the space of all real-valued tensors  $Y = (y_{i_1, \dots, i_p})_{1 \leq i_1, \dots, i_p \leq N}$ . For any two tensors  $Y$  and  $Y'$ , their outer product and inner

product are defined respectively by

$$(Y \otimes Y')_{i_1, \dots, i_p, i'_1, \dots, i'_p} = y_{i_1, \dots, i_p} y'_{i'_1, \dots, i'_p}$$

and

$$\langle Y, Y' \rangle := \sum_{i_1, \dots, i_p=1}^N y_{i_1, \dots, i_p} y'_{i_1, \dots, i_p}.$$

For  $\tau \in \mathbb{R}^N$ , we define  $\tau^{\otimes p} = \tau \otimes \dots \otimes \tau \in \Omega_N$  as the  $p$ th order power of  $\tau$ . For any permutation  $\pi \in \mathfrak{S}_p$  and any tensor  $Y, Y^\pi \in \Omega_N$  is defined as  $y_{i_1, \dots, i_p}^\pi = y_{i_{\pi(1)}, \dots, i_{\pi(p)}}$ . We say that a tensor  $Y$  is symmetric if  $Y = Y^\pi$  for all  $\pi \in \mathfrak{S}_p$ . Let  $\Omega_N^s$  be the collection of all symmetric tensors. For any measurable  $A \subseteq \mathbb{R}^k$  for some  $k \geq 1$ , we use  $\mathcal{B}(A)$  to stand for the Borel  $\sigma$ -field on  $A$ .

The symmetric Gaussian random tensor is defined as follows. Denote by  $Y \in \Omega_N$  a random tensor with i.i.d. entries  $y_{i_1, \dots, i_p} \sim N(0, 2/N)$ . Define a symmetric random tensor  $W$  by

$$W = \frac{1}{p!} \sum_{\pi \in \mathfrak{S}_p} Y^\pi.$$

For example, if  $p = 2$ ,  $W$  is the Gaussian orthogonal ensemble, that is,  $w_{ij} = w_{ji}$  are independent of each other with  $w_{ij} \sim N(0, 1/N)$  for  $i < j$  and  $w_{ii} \sim N(0, 2/N)$ . Let  $S_N := \{\pm 1/\sqrt{N}\}^N$ . Assume that  $u$  is sampled uniformly at random from  $S_N$  and is independent of  $W$ . Set the spiked random tensor as

$$T = W + \beta u^{\otimes p}$$

for  $\beta \geq 0$ . Let

$$d_{\text{TV}}(W, T) := \sup_{A \in \mathcal{B}(\Omega_N^s)} |\mathbb{P}(W \in A) - \mathbb{P}(T \in A)|$$

be the total variation distance between  $W$  and  $T$ . We now make the distinguishability and indistinguishability between  $W$  and  $T$  precise.

**DEFINITION 2.1.** We say that:

- (i)  $W$  and  $T$  are indistinguishable if  $\lim_{N \rightarrow \infty} d_{\text{TV}}(W, T) = 0$ .
- (ii)  $W$  and  $T$  are distinguishable if  $\lim_{N \rightarrow \infty} d_{\text{TV}}(W, T) = 1$ .

Let  $u$  be a realization of the prior. For a given tensor  $X \in \Omega_N^s$ , consider the detection problem that under the null hypothesis,  $X = W$  and under the alternative hypothesis,  $X = T$ . Item (i) essentially says that any statistical test can not reliably distinguish these two hypotheses. Item (ii) means there exists a sequences of events that distinguishes these two tensors. Next, we define the notion of weak recovery for  $u$ .

DEFINITION 2.2. For  $\beta > 0$ , we say that weak recovery of  $u$  is possible if there exists a sequence of random probability measures  $\mu_N$  on  $\Omega_N^s \times \mathcal{B}(S_N)$  and a constant  $c > 0$  such that

$$(2.1) \quad \lim_{N \rightarrow \infty} \mathbb{P} \left( \int_{S_N} |\langle u, \tau \rangle| \mu_N(T, d\tau) \geq c \right) = 1$$

and that weak recovery of  $u$  is not possible if for any random probability measure  $\mu_N$  on  $\Omega_N^s \times \mathcal{B}(S_N)$  and constant  $c > 0$ ,

$$(2.2) \quad \lim_{N \rightarrow \infty} \mathbb{P} \left( \int_{S_N} |\langle u, \tau \rangle| \mu_N(T, d\tau) \geq c \right) = 0.$$

Here,  $\mu_N$  is a random probability measure on  $\Omega_N^s \times \mathcal{B}(S_N)$  means that  $\mu_N$  is a mapping from  $\Omega_N^s \times \mathcal{B}(S_N)$  to  $[0, 1]$  such that  $\mu_N(\cdot, A)$  is  $\mathcal{B}(\Omega_N^s)$ -measurable for each  $A \in \mathcal{B}(S_N)$  and  $\mu_N(w, \cdot)$  is a probability measure on  $(S_N, \mathcal{B}(S_N))$  for each  $w \in \Omega_N^s$ .

A few comments are in order. Consider a given realization of signal  $u$  and tensor  $T$ . Equation (2.1) ensures that there exists some  $\tau$  produced though the measure  $\mu_N(T, d\tau)$  such that  $u$  and  $\tau$  have a nontrivial overlap. To understand (2.2), let  $\phi : \Omega_N^s \rightarrow S_N$  be any measurable function. If we consider the random probability measure  $\mu_N(w, \tau)$  defined by

$$\mu_N(w, \tau) = \begin{cases} 1, & \text{if } \tau = \phi(w), \\ 0, & \text{if } \tau \neq \phi(w), \end{cases}$$

then from (2.2), for any  $c > 0$ ,

$$\lim_{N \rightarrow \infty} \mathbb{P} (|\langle u, \phi(T) \rangle| \geq c) = 0.$$

In other words, any vector generated by  $T$  is uncorrelated with the signal  $u$ , and thus, it does not provide indicative information about  $u$ . We emphasize that Definitions 2.1 and 2.2 are not directly related to each other. Nevertheless, we will show that both of them hold up to a critical threshold in our main result below.

Now we introduce the pure  $p$ -spin model. For each  $N \geq 1$ , set  $\Sigma_N := \{\pm 1\}^N$ . The Hamiltonian of the pure  $p$ -spin model is defined by

$$H_N(\sigma) = \frac{1}{N^{(p-1)/2}} \sum_{1 \leq i_1, \dots, i_p \leq N} g_{i_1, \dots, i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

for  $\sigma \in \Sigma_N$ . Its covariance can be computed as

$$\mathbb{E} H_N(\sigma^1) H_N(\sigma^2) = N(R_{1,2})^p,$$

where  $R_{1,2}$  is the overlap between  $\sigma^1, \sigma^2 \in \Sigma_N$ ,

$$(2.3) \quad R_{1,2} := \frac{1}{N} \sum_{i=1}^N \sigma_i^1 \sigma_i^2.$$

In the terminology of detection problems,  $\beta$  is understood as the signal-to-noise ratio. In the pure  $p$ -spin model, we regard  $\beta$  as a (inverse) temperature parameter. For a given temperature  $\beta \geq 0$ , define the free energy by

$$(2.4) \quad F_N(\beta) = \frac{1}{N} \log \sum_{\sigma \in \Sigma_N} \frac{1}{2^N} \exp \frac{\beta}{\sqrt{2}} H_N(\sigma).$$

If  $p = 2$ , this model is known as the Sherrington–Kirkpatrick model and it has been intensively studied over the past decades. The reader is referred to check the books [37, 46, 47] for recent mathematical advances for the SK model as well as more general models involving a mixture of pure  $p$ -spin interactions. In particular, it is already known that the thermodynamic limit of  $F_N$  ( $N \rightarrow \infty$ ) converges to a nonrandom quantity that can be expressed as the famous Parisi formula; see, for example, [38, 45]. Denote this limit by  $F$ . A direct application of Jensen's inequality to (2.4) implies that for all  $\beta \geq 0$ ,

$$F(\beta) \leq \frac{\beta^2}{4}.$$

Define the high temperature regime for the pure  $p$ -spin model as

$$\mathcal{R} = \left\{ \beta > 0 : F(\beta) = \frac{\beta^2}{4} \right\}.$$

Set the critical temperature by

$$\beta_p := \max \mathcal{R}.$$

Our main result shows that  $\beta_p$  is the critical threshold in the detection problem.

**THEOREM 2.1.** *Let  $p \geq 3$ . The following statements hold:*

- (i) *If  $0 < \beta < \beta_p$ , then  $W$  and  $T$  are indistinguishable and weak recovery of  $u$  is impossible.*
- (ii) *If  $\beta > \beta_p$ , then  $W$  and  $T$  are distinguishable and weak recovery of  $u$  is possible.*

Our main contribution in Theorem 2.1 is the part on the indistinguishability of  $W$  and  $T$  in the statement (i). Previous results along this line were established in [42], where the authors showed that there exists some  $\beta'_p < \beta_p$  so that  $W$  and  $T$  are indistinguishable for any  $\beta < \beta'_p$ . Theorem 2.1(i) here proves that this behavior is indeed valid up to the critical value  $\beta_p$ . As one will see from Theorem 4.1 below, we give a characterization of the high temperature regime  $\mathcal{R}$  and provide one way to compute numerically  $\beta_p$  in terms of an auxiliary function deduced from the optimality of the Parisi formula for the free energy at high temperature. Indeed,  $\beta_p$  is the largest  $\beta$  such that the following inequality is valid:

$$\sup_{r \in (0, 1]} \int_0^r s^{p-2} (\rho_\beta(s) - s) ds \leq 0,$$

where for  $g$  a standard normal random variable,

$$\rho_\beta(s) := \mathbb{E} \tanh^2(\beta g \sqrt{ps^{p-1}/2}) \cosh(\beta g \sqrt{ps^{p-1}/2}) e^{-\frac{p\beta^2 s^{p-1}}{4}} \quad \forall s \in [0, 1].$$

Numerically, it is obtained that

$p$	$\beta_p$
3	1.535
4	1.621
5	1.647
6	1.657

This agrees with the prediction in [42], Figure 1. We comment that for Theorem 2.1(i), a polynomial rate of convergence for the total variation distance can also be obtained; see Remark 3.2 below. In comparison, we add that Theorem 2.1 is quite different from the BBP transition for  $p = 2$ ; see [3, 32, 35, 43]. In this case,  $\beta_2 = 1$  and it is known that for  $\beta > \beta_2$ , one can distinguish  $W$  and  $T$  in the sense of Definition 2.1 by using the top eigenvalue. For  $0 < \beta < \beta_2$ , it presents a weaker sense of distinguishability,  $\lim_{N \rightarrow \infty} d_{TV}(W, T) \in (0, 1)$ ; see [23] and Remark 3.1.

As mentioned before, the work [29] investigated the present detection problem for any given prior. Their results state that one can distinguish  $W$  and  $T$  by the MMSE method and weakly recover the signal through the MMSE estimator when  $\beta > \beta_p^{MMSE}$ ; if  $\beta < \beta_p^{MMSE}$ , they concluded that weak recovery of the signal is not possible. In other words, their results imply the weak recovery part of item (i) as well as the statement of item (ii). Their constant  $\beta_p^{MMSE}$  when applying to the Rademacher prior agrees with our critical value  $\beta_p$  here. Nevertheless, we emphasize that their approach and the way how the critical value  $\beta_p$  was discovered are fundamentally different from the argument we present here. As one will see, while Theorem 2.1(ii) follows directly from a relation (see Lemma 3.2) between the total variation distance and the free energies, the delicate part is Theorem 2.1(i), which is the major component of this paper.

Finally, we comment that although we only consider the Rademacher prior and this yields significant simplifications, we anticipate that the same results throughout this paper should be still true for general priors, such as priors with bounded supports. In fact, as mentioned in the Introduction, our result is based on the observation that the total variation distance between two spikes can be expressed as an integral of the distribution function of the free energy  $F_N(\beta)$ . In order to obtain eligible control, the Parisi formula for the limiting free energy  $F(\beta)$  and the Guerra–Talagrand inequality for the coupled free energy with overlap constraint are in position. Their extensions to general spin configuration spaces are known to be valid in [40], from which we believe that a similar line of the argument in the present paper together with a more delicate analysis would carry through in more general settings.

**3. Proof of Theorem 2.1.**

3.1. *Total variation distance.* In this subsection, we prepare some lemmas for Theorem 2.1. Recall that  $Y$  is a random tensor with i.i.d.  $y_{i_1, \dots, i_p} \sim N(0, 2/N)$  and  $W$  is the symmetric random tensor generated by  $Y$ . Throughout the remainder of the paper, we use  $I(S)$  to stand for an indicator function on a set  $S$ . We first establish an elementary expression of the total variation distance.

LEMMA 3.1. *Let  $U, V$  be two  $n$ -dimensional random vectors with densities  $f_U$  and  $f_V$  satisfying  $f_U(r) \neq 0$  and  $f_V(r) \neq 0$  a.e. on  $\mathbb{R}^n$ . Then*

$$d_{\text{TV}}(U, V) = \int_0^1 \mathbb{P}\left(\frac{f_U(V)}{f_V(V)} < x\right) dx = \int_0^1 \mathbb{P}\left(\frac{f_U(U)}{f_V(U)} > \frac{1}{x}\right) dx.$$

PROOF. Note that

$$d_{\text{TV}}(U, V) = \frac{1}{2} \int_{\mathbb{R}^n} |f_V(r) - f_U(r)| dr = \int_{f_U(r) \leq f_V(r)} (f_V(r) - f_U(r)) dr.$$

Using Fubini’s theorem and this equation, the first equality follows by

$$\begin{aligned} \int_0^1 \mathbb{P}\left(\frac{f_U(V)}{f_V(V)} < x\right) dx &= \int_0^1 \int_{\mathbb{R}^n} I\left(\frac{f_U(r)}{f_V(r)} < x\right) f_V(r) dr dx \\ &= \int_{\mathbb{R}^n} \int_0^1 I\left(\frac{f_U(r)}{f_V(r)} < x\right) dx f_V(r) dr \\ &= \int_{\mathbb{R}^n} I\left(\frac{f_U(r)}{f_V(r)} \leq 1\right) \left(1 - \frac{f_U(r)}{f_V(r)}\right) f_V(r) dr \\ &= \int_{f_U(r) \leq f_V(r)} (f_V(r) - f_U(r)) dr. \end{aligned}$$

To obtain the second equality, one simply exchanges the roles of  $U, V$ .  $\square$

Recall the free energy  $F_N(\beta)$  and the Rademacher prior  $u$  from Section 2. Define an auxiliary free energy of the pure  $p$ -spin model with a Curie–Weiss type interaction as

$$(3.1) \quad AF_N(\beta) = \frac{1}{N} \log \sum_{\sigma \in \Sigma_N} \frac{1}{2^N} \exp\left(\frac{\beta}{\sqrt{2}} H_N(\sigma) + \frac{\beta^2}{2} N \left(\frac{1}{N} \sum_{i=1}^N h_i \sigma_i\right)^p\right),$$

where  $h_i := \sqrt{N}u_i$ . In Section E of the Supplementary Material [18], it will be established that the limit of  $AF_N$  converges a.s. to a nonrandom quantity for any  $\beta \geq 0$ . Denote this limit by  $AF(\beta)$ . The following lemma relates the total variation distance between  $W$  and  $T$  to the free energy  $F_N$  and the auxiliary free energy  $AF_N$ .

LEMMA 3.2. For any  $\beta \geq 0$ , we have that

$$(3.2) \quad d_{TV}(W, T) = \int_0^1 \mathbb{P}\left(F_N(\beta) < \frac{\beta^2}{4} + \frac{\log x}{N}\right) dx$$

$$(3.3) \quad = \int_0^1 \mathbb{P}\left(AF_N(\beta) > \frac{\beta^2}{4} - \frac{\log x}{N}\right) dx.$$

PROOF. Note that  $W$  has density  $f_W(w) = \exp(-N\langle w, w \rangle/4)/C$  on  $\Omega_N^s$  for some normalizing constant  $C > 0$ . For any  $A \in \mathcal{B}(\Omega_N^s)$ , a change of variables gives

$$\begin{aligned} \mathbb{P}(T \in A) &= \mathbb{E}_u[\mathbb{P}_W(W \in A - \beta u^{\otimes p})] \\ &= \mathbb{E}_u\left[\int_{A - \beta u^{\otimes p}} f_W(w) dw\right] \\ &= \mathbb{E}_u\left[\int_A f_W(w - \beta u^{\otimes p}) dw\right] = \int_A \mathbb{E}_u f_W(w - \beta u^{\otimes p}) dw, \end{aligned}$$

where  $\mathbb{E}_u$  is the expectation with respect to  $u$  only and  $dw$  is the Lebesgue on  $\Omega_N^s$ . This implies that the density of  $T = W + \beta u^{\otimes p}$  is given by  $f_T(w) := \mathbb{E}_u f_W(w - \beta u^{\otimes p})$ . Now since

$$\begin{aligned} f_T(w) &= \frac{1}{C} \mathbb{E}_u \exp \frac{-N}{4} \langle w - \beta u^{\otimes p}, w - \beta u^{\otimes p} \rangle \\ &= f_W(w) \mathbb{E}_u \exp \frac{N}{4} (2\beta \langle w, u^{\otimes p} \rangle - \beta^2 \langle u^{\otimes p}, u^{\otimes p} \rangle), \end{aligned}$$

we obtain

$$(3.4) \quad \frac{f_T(w)}{f_W(w)} = \mathbb{E}_u \exp \frac{N}{4} (2\beta \langle w, u^{\otimes p} \rangle - \beta^2 \langle u^{\otimes p}, u^{\otimes p} \rangle).$$

Here, since  $\langle W, \tau^{\otimes p} \rangle = \langle Y, \tau^{\otimes p} \rangle$  for any  $\tau \in \mathbb{R}^N$  and  $\langle u^{\otimes p}, u^{\otimes p} \rangle = 1$ , we see that

$$\begin{aligned} \log \frac{f_T(W)}{f_W(W)} &= \log \mathbb{E}_u \exp \frac{N}{4} (2\beta \langle Y, u^{\otimes p} \rangle - \beta^2) \\ &\stackrel{d}{=} -\frac{\beta^2 N}{4} + N F_N(\beta) \end{aligned}$$

and

$$\begin{aligned} \log \frac{f_T(T)}{f_W(T)} &= \log \mathbb{E}_{u'} \exp \frac{N}{4} (2\beta \langle Y + \beta u^{\otimes p}, u'^{\otimes p} \rangle - \beta^2) \\ &\stackrel{d}{=} -\frac{\beta^2 N}{4} + N A F_N(\beta), \end{aligned}$$

where  $\mathbb{E}_{u'}$  is the expectation of  $u'$ , an independent copy of  $u$  and independent of  $W$ , and the second equality of both displays used the assumption that

$y_{i_1, \dots, i_p} \stackrel{d}{=} g_{i_1, \dots, i_p} \sqrt{2/N}$  and  $\sqrt{N}S_N = \Sigma_N$ . Our proof is then completed by applying Lemma 3.1.  $\square$

REMARK 3.1. Aizenman, Lebowitz and Ruelle [1] showed that  $N(F_N(\beta) - \beta^2/4)$  converges to a Gaussian random variable. From (3.2), one immediately sees that  $\lim_{N \rightarrow \infty} d_{TV}(W, T) \in (0, 1)$ .

3.2. *Proof of Theorem 2.1(i).* The central ingredient throughout our proof is played by the high temperature behavior of the pure  $p$ -spin model stated in Section 4 below, namely, a tight upper bound for the fluctuation of the free energy in Proposition 4.1 and a good moment control for the concentration of the overlap  $R_{1,2}$  around zero under the Gibbs measure in Theorem 4.3. The former will be directly used to show that the total variation distance  $d_{TV}(W, T)$  vanishes via the exact expression (3.2), while the latter is vital in order to establish the impossibility of weak recovery of  $u$ .

PROOF OF THEOREM 2.1(i): INDISTINGUISHABILITY. Let  $p \geq 3$ . Assume that  $0 < \beta < \beta_p$ . For any  $0 < \varepsilon < 1$ , writing  $\int_0^1 = \int_0^{1-\varepsilon} + \int_{1-\varepsilon}^1$  in (3.2) gives

$$\begin{aligned} d_{TV}(W, T) &\leq \varepsilon + \int_0^{1-\varepsilon} \mathbb{P}\left(F_N(\beta) < \frac{\beta^2}{4} + \frac{\log x}{N}\right) dx \\ &\leq \varepsilon + \mathbb{P}\left(F_N(\beta) < \frac{\beta^2}{4} - \frac{\log(1-\varepsilon)^{-1}}{N}\right) \\ &\leq \varepsilon + \mathbb{P}\left(\left|F_N(\beta) - \frac{\beta^2}{4}\right| \geq \frac{\log(1-\varepsilon)^{-1}}{N}\right). \end{aligned}$$

To complete the proof, we use a key property about the fluctuation of the free energy stated in Proposition 4.1 below, which says that there exists a constant  $K$  such that

$$\mathbb{P}\left(\left|F_N(\beta) - \frac{\beta^2}{4}\right| \geq \frac{\log(1-\varepsilon)^{-1}}{N}\right) \leq \frac{K}{(\log(1-\varepsilon)^{-1})^2 N^{\frac{p}{2}-1}}$$

for all  $N \geq 1$ . From this,

$$(3.5) \quad d_{TV}(W, T) \leq \varepsilon + \frac{K}{(\log(1-\varepsilon)^{-1})^2 N^{\frac{p}{2}-1}}.$$

Since  $p \geq 3$ , sending  $N \rightarrow \infty$  and then  $\varepsilon \downarrow 0$  implies that  $W$  and  $T$  are indistinguishable.  $\square$

Next, we continue to show that weak recovery of  $u$  is impossible. For  $\beta > 0$ , define a random probability measure on  $\Omega_N^s \times \mathcal{B}(S_N)$  by

$$(3.6) \quad \nu_{N,\beta}(w, A) = \frac{\mathbb{E}_{u'}[\exp \frac{\beta N}{2} \langle w, u'^{\otimes p} \rangle; A]}{\mathbb{E}_{u'}[\exp \frac{\beta N}{2} \langle w, u'^{\otimes p} \rangle]}$$

for any  $w \in \Omega_N^s$  and  $A \in \mathcal{B}(S_N)$ , where  $\mathbb{E}_{u'}$  is the expectation with respect to  $u'$ , an independent copy of  $u$ . The following lemma relates the expectation of  $(u, T)$  to  $W$  by a change of measure in terms of  $\mathbb{E}_{\nu_{N,\beta}}$ .

LEMMA 3.3. *Let  $\zeta_N$  be a measurable function from  $S_N \times \Omega_N^s$  to  $[0, 1]$ . If  $W$  and  $T$  are indistinguishable, then*

$$\lim_{N \rightarrow \infty} \left| \mathbb{E} \zeta_N(u, T) - \mathbb{E} \int_{S_N} \zeta_N(\tau, W) \nu_{N,\beta}(W, d\tau) \right| = 0.$$

PROOF. Recall the densities  $f_W$  and  $f_T$  of  $W$  and  $T$  from Lemma 3.2. Let  $f_u(\tau)$  be the probability mass function for  $u$ . Since  $u$  is independent of  $W$ , the joint density of  $(u, T)$  is given by

$$f_u(\tau) f_W(w - \beta \tau^{\otimes p}).$$

This implies that  $\mathbb{E}[\zeta_N(u, T) | T] = \zeta_N(T)$ , where

$$\zeta_N(w) := \frac{\sum_{\tau \in S_N} \zeta_N(\tau, w) f_u(\tau) f_W(w - \beta \tau^{\otimes p})}{f_T(w)}.$$

Note that  $0 \leq \zeta_N(w) \leq 1$  since  $0 \leq \zeta_N(\tau, w) \leq 1$ . For any  $k \geq 1$ , define

$$\phi_k(s) = \begin{cases} \frac{i}{k}, & \text{if } s \in A_{k,i} \text{ for some } 1 \leq i \leq k - 1, \\ 1, & \text{if } s \in A_{k,k}, \end{cases}$$

where  $A_{k,i} := [(i - 1)/k, i/k]$  for  $1 \leq i \leq k - 1$  and  $A_{k,k} := [(k - 1)/k, 1]$ . Observe that  $|\phi_k(s) - s| \leq 1/k$  for  $s \in [0, 1]$ . From this and the triangle inequality,

$$\begin{aligned} & |\mathbb{E} \zeta_N(T) - \mathbb{E} \zeta_N(W)| \\ & \leq |\mathbb{E} \zeta_N(T) - \mathbb{E} \phi_k(\zeta_N(T))| \\ & \quad + |\mathbb{E} \phi_k(\zeta_N(T)) - \mathbb{E} \phi_k(\zeta_N(W))| + |\mathbb{E} \zeta_N(W) - \mathbb{E} \phi_k(\zeta_N(W))| \\ & \leq \frac{2}{k} + \sum_{i=1}^k \frac{i}{k} |\mathbb{P}(\zeta_N(T) \in A_{k,i}) - \mathbb{P}(\zeta_N(W) \in A_{k,i})|. \end{aligned}$$

Since  $d_{TV}(W, T)$  converges to zero, each term in the above sum must vanish in the limit and thus, letting  $N \rightarrow \infty$  and then  $k \rightarrow \infty$  yields

$$(3.7) \quad \lim_{N \rightarrow \infty} |\mathbb{E} \zeta_N(T) - \mathbb{E} \zeta_N(W)| = 0.$$

Now write

$$\mathbb{E} \zeta_N(W) = \int_{\Omega_N^s} \sum_{\tau \in S_N} \zeta_N(\tau, w) \frac{f_u(\tau) f_W(w - \beta \tau^{\otimes p}) f_W(w)}{f_T(w)} dw.$$

Note that  $f_W(w) = \exp(-N\langle w, w \rangle/4)/C$  for some normalizing constant. Since from (3.4) and  $\langle \tau^{\otimes p}, \tau^{\otimes p} \rangle = \langle u'^{\otimes p}, u'^{\otimes p} \rangle = 1$ ,

$$\begin{aligned} & \frac{f_u(\tau) f_W(w - \beta\tau^{\otimes p}) f_W(w)}{f_T(w)} \\ &= \frac{\exp \frac{N}{4} (-\langle w, w \rangle + 2\beta \langle w, \tau^{\otimes p} \rangle - \beta^2 \langle \tau^{\otimes p}, \tau^{\otimes p} \rangle)}{C 2^N \mathbb{E}_{u'} [\exp \frac{N}{4} (2\beta \langle w, u'^{\otimes p} \rangle - \beta^2 \langle u'^{\otimes p}, u'^{\otimes p} \rangle)]} \\ &= f_W(w) \nu_{N,\beta}(w, \tau), \end{aligned}$$

it follows that

$$\mathbb{E} \zeta_N(W) = \mathbb{E} \int_{S_N} \zeta_N(\tau, W) \nu_{N,\beta}(W, d\tau).$$

From this and (3.7), the announced result follows.  $\square$

**PROOF OF THEOREM 2.1(i): IMPOSSIBILITY OF WEAK RECOVERY.** Let  $p \geq 3$  and  $0 < \beta < \beta_p$ . Let  $\mu_N$  be a random probability measure on  $\Omega_N^S \times \mathcal{B}(S_N)$  (see Definition 2.2) and  $c > 0$ . Our goal is to show that

$$(3.8) \quad \lim_{N \rightarrow \infty} \mathbb{P} \left( \int_{S_N} |\langle u, \tau \rangle| \mu_N(T, d\tau) \geq c \right) = 0.$$

Set

$$\zeta_N(\tau, w) = \int_{S_N} \langle \tau, \tau' \rangle^2 \mu_N(w, d\tau')$$

for  $(\tau, w) \in S_N \times \Omega_N^S$ . Note that  $\zeta_N \in [0, 1]$  and  $\zeta_N$  is measurable. From Lemma 3.3,

$$(3.9) \quad \begin{aligned} & \lim_{N \rightarrow \infty} \left| \mathbb{E} \int_{S_N} \langle u, \tau' \rangle^2 \mu_N(T, d\tau') \right. \\ & \left. - \mathbb{E} \int_{S_N \times S_N} \langle \tau, \tau' \rangle^2 \mu_N(W, d\tau') \nu_{N,\beta}(W, d\tau) \right| = 0. \end{aligned}$$

We claim that the second expectation converges to zero. For notation convenience, we simply denote  $\mu_N(d\tau') = \mu_N(W, d\tau')$  and  $\nu_N(d\tau) = \nu_{N,\beta}(W, d\tau)$ . Note that

$$\langle \tau, \tau' \rangle^2 = \sum_{i_1, i_2=1}^N \tau_{i_1} \tau_{i_2} \tau'_{i_1} \tau'_{i_2}.$$

The second term in the above equation can be controlled by

$$\mathbb{E} \int_{S_N \times S_N} \langle \tau, \tau' \rangle^2 \mu_N(d\tau') \nu_N(d\tau)$$

$$\begin{aligned}
 &= \mathbb{E} \sum_{i_1, i_2=1}^N \int_{S_N \times S_N} \tau_{i_1} \tau_{i_2} \tau'_{i_1} \tau'_{i_2} \mu_N(d\tau') \nu_N(d\tau) \\
 &= \sum_{i_1, i_2=1}^N \mathbb{E} \int_{S_N} \tau_{i_1} \tau_{i_2} \nu_N(d\tau) \cdot \int_{S_N} \tau'_{i_1} \tau'_{i_2} \mu_N(d\tau') \\
 &\leq \sum_{i_1, i_2=1}^N \left( \mathbb{E} \left( \int_{S_N} \tau_{i_1} \tau_{i_2} \nu_N(d\tau) \right)^2 \right)^{1/2} \cdot \left( \mathbb{E} \left( \int_{S_N} \tau'_{i_1} \tau'_{i_2} \mu_N(d\tau') \right)^2 \right)^{1/2},
 \end{aligned}$$

where the last inequality used the Cauchy–Schwarz inequality. Using the Cauchy–Schwarz inequality again, the last inequality is bounded above by

$$\begin{aligned}
 &\left( \sum_{i_1, i_2=1}^N \mathbb{E} \left( \int_{S_N} \tau_{i_1} \tau_{i_2} \nu_N(d\tau) \right)^2 \right)^{1/2} \cdot \left( \sum_{i_1, i_2=1}^N \mathbb{E} \left( \int_{S_N} \tau'_{i_1} \tau'_{i_2} \mu_N(d\tau') \right)^2 \right)^{1/2} \\
 &= \left( \sum_{i_1, i_2=1}^N \mathbb{E} \int_{S_N \times S_N} \tau_{i_1} \tau_{i_2} \hat{\tau}_{i_1} \hat{\tau}_{i_2} \nu_N(d\tau) \nu_N(d\hat{\tau}) \right)^{1/2} \\
 &\quad \cdot \left( \sum_{i_1, i_2=1}^N \mathbb{E} \int_{S_N \times S_N} \tau'_{i_1} \tau'_{i_2} \hat{\tau}'_{i_1} \hat{\tau}'_{i_2} \mu_N(d\tau') \mu_N(d\hat{\tau}') \right)^{1/2} \\
 &= \left( \mathbb{E} \int_{S_N \times S_N} \langle \tau, \hat{\tau} \rangle^2 \nu_N(d\tau) \nu_N(d\hat{\tau}) \right)^{1/2} \\
 &\quad \cdot \left( \mathbb{E} \int_{S_N \times S_N} \langle \tau', \hat{\tau}' \rangle^2 \mu_N(d\tau') \mu_N(d\hat{\tau}') \right)^{1/2}.
 \end{aligned}$$

Here, the second bracket is bounded above by 1. As for the first one, we observe that  $\nu_N$  is in distribution equal to the Gibbs measure  $G_{N,\beta}$  defined in (4.2) and if we write  $\sigma^1 = \sqrt{N}\tau$  and  $\sigma^2 = \sqrt{N}\tau'$ , then in distribution,  $\sigma^1, \sigma^2$  are independent samplings from  $G_{N,\beta}$  and  $\langle \tau, \hat{\tau} \rangle$  is the overlap  $R_{1,2}$  between  $\sigma^1$  and  $\sigma^2$ . As a result,

$$\mathbb{E} \int_{S_N \times S_N} \langle \tau, \hat{\tau} \rangle^2 \nu_N(d\tau) \nu_N(d\hat{\tau}) = \mathbb{E} \langle R_{1,2}^2 \rangle_\beta,$$

where  $\langle \cdot \rangle_\beta$  is the Gibbs average with respect to the product measure  $G_{N,\beta} \times G_{N,\beta}$ . Now, since  $0 < \beta < \beta_p$ , we can apply Theorem 4.3 to control the right-hand side by the bound

$$\mathbb{E} \langle R_{1,2}^2 \rangle_\beta \leq \frac{K}{N}$$

for some constant  $K$  independent of  $N$ . Hence, from the above inequalities,

$$\lim_{N \rightarrow \infty} \mathbb{E} \int_{S_N \times S_N} \langle \tau, \tau' \rangle^2 \mu_N(d\tau') \nu_N(d\tau) = 0.$$

From (3.9),

$$\lim_{N \rightarrow \infty} \mathbb{E} \int_{S_N} \langle u, \tau' \rangle^2 \mu_N(T, d\tau') = 0,$$

which gives the desired limit (3.8) by using Markov’s and Jensen’s inequalities.  $\square$

REMARK 3.2. Take  $\varepsilon = N^{-\delta}$  for  $\delta = (p/2 - 1)/3$  and use  $\log(1 - \varepsilon)^{-1} \geq \varepsilon$  in (3.5). We obtain the rate of convergence,

$$d_{\text{TV}}(W, T) \leq \varepsilon + \frac{K}{\varepsilon^2 N^{\frac{p}{2}-1}} = \frac{1}{N^\delta} + \frac{K}{N^{\frac{p}{2}-1-2\delta}} = \frac{1+K}{N^{\frac{1}{3}(\frac{p}{2}-1)}}.$$

3.3. *Proof of Theorem 2.1(ii).* While we have seen that the high temperature behavior of the pure  $p$ -spin model has been of great use in obtaining Theorem 2.1(i), the proof of Theorem 2.1(ii) below relies only on the low temperature behavior of the free energies, that is,  $F(\beta) < \beta^2/4$  and  $AF(\beta) > \beta^2/4$  for  $\beta > \beta_p$ . The proof is relatively simpler than that for Theorem 2.1(i).

PROOF OF THEOREM 2.1(ii): DISTINGUISHABILITY. Let  $p \geq 3$ . Assume that  $\beta > \beta_p$ . Since  $F(\beta) < \beta^2/4$  and  $F_N(\beta) - \log x/N$  converges to  $F(\beta)$  a.s.,

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(F_N(\beta) < \frac{\beta^2}{4} + \frac{\log x}{N}\right) = \mathbb{P}\left(F(\beta) \leq \frac{\beta^2}{4}\right) = 1.$$

Thus, from (3.2) and the dominated convergence theorem,  $W$  and  $T$  are distinguishable.  $\square$

Next, we show that weak recovery of  $u$  is possible. Recall  $h_i$  from the definition of  $AF_N(\beta)$ . Let  $u'$  be an independent copy of  $u$  and be independent of  $Y$ . For fixed  $\beta > 0$ , define an interpolating free energy between  $F_N(\beta)$  and  $AF_N(\beta)$  by

$$\begin{aligned} L_N(x) &= \frac{1}{N} \log \mathbb{E}_{u'} \exp \frac{\beta N}{2} \langle Y + xu^{\otimes p}, u'^{\otimes p} \rangle \\ (3.10) \quad &\stackrel{d}{=} \frac{1}{N} \log \sum_{\sigma \in \Sigma_N} \frac{1}{2^N} \exp \left( \frac{\beta}{\sqrt{2}} H_N(\sigma) + \frac{\beta x}{2} \left( \frac{1}{N} \sum_{i=1}^N h_i \sigma_i \right)^p \right) \end{aligned}$$

for  $x > 0$ , where  $\mathbb{E}_{u'}$  is the expectation with respect to  $u'$  only. Note that in distribution  $L_N(0) = F_N(\beta)$  and  $L_N(\beta) = AF_N(\beta)$ . Similar to  $AF_N(\beta)$ , it can be shown that  $L_N$  also converges to a nonrandom quantity for any  $x > 0$ ; see Proposition E.1 in Section E in of the Supplementary Material [18]. Denote this limit by  $L$ . From now on, we use  $D_+ f$  and  $D_- f$  to denote the right and left derivatives of  $f$  whenever they exist. Note that since  $L$  is convex,  $D_- L$  exists everywhere. Recall  $\nu_{N,\beta}$  from (3.6).

LEMMA 3.4. *Let  $\beta > 0$ . For any  $\varepsilon > 0$ , we have*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \int_{S_N} |\langle u, \tau \rangle|^p \nu_{N,\beta}(T, d\tau) \geq D-L(\beta) - \varepsilon \right) = 1.$$

PROOF. Note that  $\langle u^{\otimes p}, u'^{\otimes p} \rangle = \langle u, u' \rangle^p$ . For any  $\eta > 0$ ,

$$\begin{aligned} & \frac{1}{N} \log \mathbb{E}_{u'} \left[ \exp \frac{\beta N}{2} \langle Y + \beta u^{\otimes p}, u'^{\otimes p} \rangle; \langle u, u' \rangle^p \leq D-L(\beta) - \varepsilon \right] \\ & \leq \frac{1}{N} \log \mathbb{E}_{u'} \left[ \exp \frac{\beta N}{2} (\langle Y, u'^{\otimes p} \rangle + (\beta - \eta) \langle u^{\otimes p}, u'^{\otimes p} \rangle) \right] + \eta(D-L(\beta) - \varepsilon) \\ & = L_N(\beta - \eta) + \eta(D-L(\beta) - \varepsilon). \end{aligned}$$

To control the last inequality, write

$$\begin{aligned} L_N(\beta - \eta) + \eta(D-L(\beta) - \varepsilon) &= \eta \left( D-L(\beta) - \frac{L_N(\beta) - L_N(\beta - \eta)}{\eta} \right) \\ & \quad + (L_N(\beta) - \eta\varepsilon) \end{aligned}$$

and pass to limit

$$\begin{aligned} \lim_{N \rightarrow \infty} (L_N(\beta - \eta) + \eta(D-L(\beta) - \varepsilon)) &= \eta \left( D-L(\beta) - \frac{L(\beta) - L(\beta - \eta)}{\eta} \right) \\ & \quad + (L(\beta) - \eta\varepsilon). \end{aligned}$$

Here, by the left-differentiability of  $L$ , the first bracket on the right-hand side converges to zero as  $\eta \downarrow 0$ . Thus, we can choose  $\eta$  small enough such that the right-hand side is controlled by  $L(\beta) - \eta\varepsilon/2$ . Consequently, from the sub-Gaussian concentration inequality for  $L_N$  (see [16], Proposition 9), we see that there exists a positive constant  $K$  independent of  $N$  such that with probability at least  $1 - Ke^{-N/K}$ ,

$$\frac{1}{N} \log \mathbb{E}_{u'} \left[ \exp \frac{\beta N}{2} \langle T, u'^{\otimes p} \rangle; \langle u, u' \rangle^p \leq D-L(\beta) - \varepsilon \right] \leq L_N(\beta) - \frac{\eta\varepsilon}{4}$$

and this implies that

$$\nu_{N,\beta}(T, \{ \tau \in S_N | \langle u, \tau \rangle^p \leq D-L(\beta) - \varepsilon \}) \leq e^{-\frac{\eta\varepsilon N}{4}}.$$

As a result, the assertion follows by

$$\begin{aligned} & \mathbb{P} \left( \int_{S_N} \langle u, \tau \rangle^p \nu_{N,\beta}(T, d\tau) \geq (D-L(\beta) - \varepsilon)(1 - e^{-\frac{\eta\varepsilon N}{4}}) - e^{-\frac{\eta\varepsilon N}{4}} \right) \\ & \geq 1 - Ke^{-\frac{N}{K}} \end{aligned}$$

and noting that  $\int_{S_N} \langle u, \tau \rangle^p \nu_{N,\beta}(T, d\tau) \leq \int_{S_N} |\langle u, \tau \rangle|^p \nu_{N,\beta}(T, d\tau)$ .  $\square$

PROOF OF THEOREM 2.1(ii): POSSIBILITY OF WEAK RECOVERY. Let  $p \geq 3$ . Assume that  $\beta > \beta_p$ . First, we claim that

$$(3.11) \quad AF(\beta) \geq \frac{\beta^2}{4}.$$

Assume on the contrary that  $AF(\beta) < \beta^2/4$ . Since  $AF_N(\beta) + \log x/N$  converges to  $AF(\beta)$  a.s., we have

$$\mathbb{P}\left( AF_N(\beta) > \frac{\beta^2}{4} - \frac{\log x}{N} \right) \rightarrow \mathbb{P}\left( AF(\beta) > \frac{\beta^2}{4} \right) = 0.$$

This and (3.3) together leads to a contradiction,

$$1 = \lim_{N \rightarrow \infty} d_{TV}(W, T) = \int_0^1 \lim_{N \rightarrow \infty} \mathbb{P}\left( AF_N(\beta) > \frac{\beta^2}{4} - \frac{\log x}{N} \right) dx = 0.$$

Thus, (3.11) must be valid.

To show that weak recovery is possible, observe that since  $L_N(\beta) = AF_N(\beta)$  and  $L_N(0) = F_N(\beta)$  in distribution, it follows from (3.11) that

$$L(\beta) = AF(\beta) \geq \frac{\beta^2}{4} > F(\beta) = L(0).$$

Since  $L$  is convex, there exists a point  $x_0 \in (0, \beta)$  such that  $D_-L(x_0) > 0$ . Indeed, if not  $L$  will be a constant function on  $(0, \beta)$ , a contradiction. Now using the convexity of  $L$  again gives  $D_-L(\beta) \geq D_-L(x_0) > 0$ . This and Lemma 3.4 together complete our proof by letting  $\mu_N(w, \tau) = \nu_{N,\beta}(w, \tau)$  and  $c = D_-L(\beta)/2$  and noting that  $\int_{S_N} |\langle u, \tau \rangle|^p \nu_{N,\beta}(T, d\tau) \leq \int_{S_N} |\langle u, \tau \rangle| \nu_{N,\beta}(T, d\tau)$ .  $\square$

**4. The pure  $p$ -spin model.** Recall the pure  $p$ -spin Hamiltonian  $H_N$  and the high temperature regime  $\mathcal{R}$  from Section 2. The aim of this section is to establish a complete description of the high temperature behavior of the pure  $p$ -spin model.

4.1. *High temperature behavior.* As mentioned before, the limiting free energy  $F_N(\beta)$  converges to a nonrandom quantity  $F(\beta)$ . This quantity can also be expressed in terms of the famous Parisi formula, which we state as follows. Throughout the rest of the paper, we set

$$\xi(s) = \frac{s^p}{2}$$

for  $s \in [0, 1]$ . Let  $\mathcal{M}$  be the collection of all cumulative distribution functions on  $[0, 1]$  equipped with the  $L^1$  distance with respect to the Lebesgue measure. This is usually called the space of functional order parameters in physics. For  $\beta > 0$ , define a functional  $\mathcal{P}_\beta$  on  $\mathcal{M}$  by

$$\mathcal{P}_\beta(\alpha) = \Phi_{\beta,\alpha}(0, 0) - \frac{\beta^2}{2} \int_0^1 \alpha(s) \xi''(s) s ds,$$

where  $\Phi_{\beta,\alpha}$  is the weak solution [24] to the following PDE:

$$\partial_t \Phi_{\beta,\alpha} = -\frac{\beta^2 \xi''}{2} (\partial_{xx} \Phi_{\beta,\alpha} + \alpha (\partial_x \Phi_{\beta,\alpha})^2)$$

with boundary condition  $\Phi_{\beta,\alpha}(1, x) = \log \cosh x$ . For any  $\beta > 0$ , the Parisi formula states that

$$\lim_{N \rightarrow \infty} F_N(\beta) = \inf_{\alpha \in \mathcal{M}} \mathcal{P}_\beta(\alpha).$$

Although we only consider the pure  $p$ -spin model here, this formula also holds in more general setting. Indeed, Talagrand [45] established the Parisi formula in the case of the mixed even  $p$ -spin models. Later Panchenko [38] extends its validity to general mixtures of the model. Recently, it was understood by Auffinger and Chen [2] that the functional  $\mathcal{P}_\beta$  is strictly convex, which guarantees the uniqueness of the minimizer for  $\mathcal{P}_\beta$ . We shall call this minimizer the Parisi measure and denote it by  $\alpha_p$ .

Recall that the high temperature regime  $\mathcal{R}$  of  $H_N$  is the collection of all  $\beta > 0$  that satisfy  $F(\beta) = \beta^2/4$ . By a direct computation, the validity of this equation is the same as saying that the Parisi measure satisfies  $\alpha_p \equiv 1$ , concluding from the uniqueness of the Parisi measure. This case is usually called the *replica symmetric* solution of the model in physics literature [30]. Recall that the critical temperature  $\beta_p$  is defined as the maximum of  $\mathcal{R}$ . Let  $g$  be a standard normal random variable. For  $\beta > 0$ , define an auxiliary function by

$$\rho_\beta(s) := \mathbb{E} \tanh^2(\beta g \sqrt{\xi'(s)}) \cosh(\beta g \sqrt{\xi'(s)}) e^{-\frac{\beta^2}{2} \xi'(s)} \quad \forall s \in [0, 1].$$

Our first theorem provides one way to characterize  $\mathcal{R}$  and  $\beta_p$ .

**THEOREM 4.1.** *For any  $p \geq 2$ ,  $\mathcal{R} = (0, \beta_p]$ . In addition, the following two statements hold:*

(i) *Let  $\beta > 0$ . Then  $\beta \in \mathcal{R}$  if and only if*

$$\int_0^r \xi''(s)(\rho_\beta(s) - s) ds \leq 0 \quad \forall r \in (0, 1].$$

(ii) *If  $0 < \beta < \beta_p$ , then*

$$(4.1) \quad \int_0^r \xi''(s)(\rho_\beta(s) - s) ds < 0 \quad \forall r \in (0, 1].$$

Item (i) is essentially the first-order optimality condition in order to obtain the replica symmetric solution. Item (ii) states that the replica symmetric solution is stable if  $\beta$  stays away from the criticality. This is the most crucial property that will allow us to establish the desired high temperature behavior of the overlap all the way up to the critical temperature.

Define the Gibbs measure by

$$(4.2) \quad G_{N,\beta}(\sigma) = \frac{\exp \beta H_N(\sigma)}{\sum_{\sigma' \in \Sigma_N} \exp \beta H_N(\sigma')}.$$

For i.i.d. samplings  $\sigma^1, \sigma^2$  from  $G_{N,\beta}$ , we use  $\langle \cdot \rangle_\beta$  to denote the expectation with respect to the product measure  $G_{N,\beta}^{\otimes 2}$ . Recall the overlap  $R_{1,2}$  between  $\sigma^1$  and  $\sigma^2$  from (2.3). Our next two theorems show that the overlap is concentrated around 0 in the high temperature regime with exponential tail probability and moment control.

**THEOREM 4.2.** *Assume that  $p \geq 2$ . Fix  $0 < a < b < \beta_p$ . For any  $\varepsilon > 0$ , there exists a constant  $K$  such that for any  $\beta \in [a, b]$ ,*

$$(4.3) \quad \mathbb{E} \langle I(|R_{1,2}| \geq \varepsilon) \rangle_\beta \leq K \exp\left(-\frac{N}{K}\right) \quad \forall N \geq 1,$$

where  $I$  is an indicator function.

**THEOREM 4.3.** *Assume that  $p \geq 3$ . Fix  $0 < b < \beta_p$ . For any  $k \geq 1$ , there exists a constant  $K > 0$  such that for any  $\beta \in [0, b]$ ,*

$$\mathbb{E} \langle R_{1,2}^{2k} \rangle_\beta \leq \frac{K}{N^k} \quad \forall N \geq 1.$$

In the case of the Sherrington–Kirkpatrick model ( $p = 2$ ), it was computed that  $\beta_2 = 1$  (see Remark 2 in [17]) and the same results as Theorems 4.2 and 4.3 were obtained in Talagrand’s book [47], Chapters 11 and 13. As for  $p \geq 3$ , Bardina, Márquez, Rovira and Tindel [10] established Theorem 4.3 for some  $b \ll \beta_p$  as  $p$  increases. Our main contribution here is that the concentration of the overlap is valid up to the critical temperature. As an application of Theorem 4.3, we deduce a control on the fluctuation of the free energy in high temperature regime.

**PROPOSITION 4.1.** *Assume that  $p \geq 3$ . Fix  $0 < b < \beta_p$ . There exists a constant  $K$  such that for any  $0 \leq \beta \leq b$ ,*

$$(4.4) \quad \mathbb{P} \left( \left| F_N(\beta) - \frac{\beta^2}{4} \right| \geq r \right) \leq \frac{K}{r^2 N^{p/2+1}}$$

for any  $r > 0$  and  $N \geq 1$ .

This theorem basically says that the fluctuation of  $F_N(\beta)$  is at most of the order  $N^{-p/4-1/2}$ . Indeed, if there exists some  $\delta_N \uparrow \infty$  such that

$$\mathbb{P} \left( \left| F_N(\beta) - \frac{\beta^2}{4} \right| \geq \delta_N N^{-\frac{p}{4}-\frac{1}{2}} \right) \geq c > 0$$

for all  $N \geq 1$ , then this contradicts (4.4). For  $p = 2$ , Aizenman, Lebowitz and Ruelle [1] proved that  $N^{-p/4-1/2} = N^{-1}$  is the right order of the fluctuation for  $F_N(\beta)$  and  $N(F_N(\beta) - \beta^2/4)$  converges to a Gaussian random variable up to the critical temperature  $\beta_p = 1$ . Similarly, for even  $p \geq 4$ , Bovier, Kurkova and Löwe [14] also showed that  $N^{p/4+1/2}(F_N(\beta) - \beta^2/4)$  has a Gaussian fluctuation up to certain temperature strictly less than  $\beta_p$ . From these, it is tempting to conjecture that  $N^{p/4+1/2}(F_N(\beta) - \beta^2/4)$  follows Gaussian law in the weak limit in the entire high temperature regime for all  $p \geq 3$ . Based on Theorems 4.2 and 4.3, this should be achievable by an adoption of the argument for the Sherrington–Kirkpatrick model [47], Section 11.4. We do not pursue this direction here.

REMARK 4.1. One can also consider the mixed  $p$ -spin model, that is, the Hamiltonian  $H_N$  is again a Gaussian process on  $\Sigma_N$  with zero mean and covariance structure  $\mathbb{E}H_N(\sigma^1)H_N(\sigma^2) = N\xi(R_{1,2})$  for some

$$\xi(s) := \frac{1}{2} \sum_{p \geq 2} c_p s^p$$

with  $c_p \geq 0$  and  $\sum_{p \geq 2} c_p = 1$ . In a similar manner, one can define its free energy and high temperature regime as those for  $F_N(\beta)$  and  $\mathcal{R}$ . In this general setting, it can be checked that Theorem 4.1 remains valid. As for Theorems 4.2 and 4.3 and Proposition 4.1, they also hold as long as there exists some  $p \geq 3$  such that  $c_p \neq 0$  and  $c_{p'} = 0$  for all  $2 \leq p' < p$ .

**Acknowledgments.** The author thanks G. Ben Arous for introducing this project to him and A. Auffinger and A. Jagannath for some fruitful discussions. He is indebted to D. Panchenko for asking good questions that led to a major improvement on the result of weak recovery of  $u$  and several suggestions regarding the presentation of the paper. He also thanks J. Barbier, A. Montanari and L. Zdeborová for suggesting several references and valuable comments.

## SUPPLEMENTARY MATERIAL

**Supplement to “Phase transition in the spiked random tensor with Rademacher prior”** (DOI: [10.1214/18-AOS1763SUPP](https://doi.org/10.1214/18-AOS1763SUPP); .pdf). The proofs of Theorems 4.1, 4.2, 4.3 and Proposition 4.1 are provided in detail in the Supplementary Material [18]. In addition, the convergence of the free energies  $AF_N$  and  $L_N$  defined respectively by (3.1) and (3.10) are established.

## REFERENCES

- [1] AIZENMAN, M., LEBOWITZ, J. L. and RUELLE, D. (1987). Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Comm. Math. Phys.* **112** 3–20. [MR0904135](https://doi.org/10.1007/BF01206802)
- [2] AUFFINGER, A. and CHEN, W.-K. (2015). The Parisi formula has a unique minimizer. *Comm. Math. Phys.* **335** 1429–1444. [MR3320318](https://doi.org/10.1007/s00033-015-0608-1)

- [3] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. [MR2165575](#)
- [4] BARBIER, J., DIA, M., MACRIS, N. and KRZAKALA, F. (2016). The mutual information in random linear estimation. In *54th Annual Allerton Conference on Communication, Control, and Computing* 625–632.
- [5] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Adv. Neural Inf. Process. Syst.* **29** 424–432.
- [6] BARBIER, J., KRZAKALA, F., MACRIS, N., MIOLANE, L. and ZDEBOROVÁ, L. (2017). Phase transitions, optimal errors and optimality of message-passing in generalized linear models. ArXiv e-prints.
- [7] BARBIER, J. and MACRIS, N. (2017). The stochastic interpolation method: A simple scheme to prove replica formulas in bayesian inference. ArXiv e-prints.
- [8] BARBIER, J., MACRIS, N., DIA, M. and KRZAKALA, F. (2017). Mutual information and optimality of approximate message-passing in random linear estimation. In ArXiv E-prints.
- [9] BARBIER, J., MACRIS, N. and MIOLANE, L. (2017). The layered structure of tensor estimation and its mutual information. In *55th Annual Allerton Conference on Communication, Control, and Computing*.
- [10] BARDINA, X., MÁRQUEZ-CARRERAS, D., ROVIRA, C. and TINDEL, S. (2004). The  $p$ -spin interaction model with external field. *Potential Anal.* **21** 311–362. [MR2081143](#)
- [11] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory* **57** 764–785. [MR2810285](#)
- [12] BEN AROUS, G., MEI, S., MONTANARI, A. and NICA, M. (2017). The landscape of the spiked tensor model. ArXiv e-prints.
- [13] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. [MR2782201](#)
- [14] BOVIER, A., KURKOVA, I. and LÖWE, M. (2002). Fluctuations of the free energy in the REM and the  $p$ -spin SK models. *Ann. Probab.* **30** 605–651. [MR1905853](#)
- [15] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. [MR2489158](#)
- [16] CHEN, W.-K. (2014). Chaos in the mixed even-spin models. *Comm. Math. Phys.* **328** 867–901. [MR3201215](#)
- [17] CHEN, W.-K. (2017). Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *Ann. Probab.* **45** 3929–3966. [MR3729619](#)
- [18] CHEN, W.-K. (2019). Supplement to “Phase transition in the spiked random tensor with Rademacher prior.” DOI:[10.1214/18-AOS1763SUPP](#).
- [19] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. [MR3671474](#)
- [20] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse pca. In *IEEE International Symposium on Information Theory*. 2197–2201.
- [21] DONNOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [22] DONNOHO, D. L., MALEKI, A. and MONTANARI, A. (2010). Message-passing algorithms for compressed sensing: I. motivation and construction. In *IEEE Information Theory Workshop (ITW)*. 115–144.
- [23] EL ALAOUI, A., KRZAKALA, F. and JORDAN, M. I. (2017). Finite-size corrections and likelihood ratio fluctuations in the spiked wigner model. ArXiv e-prints.

- [24] JAGANNATH, A. and TOBASCO, I. (2016). A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.* **144** 3135–3150. [MR3487243](#)
- [25] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. [MR3311445](#)
- [26] KORADA, S. B. and MACRIS, N. (2009). Exact solution of the gauge symmetric  $p$ -spin glass model on a complete graph. *J. Stat. Phys.* **136** 205–230. [MR2525244](#)
- [27] KRZAKALA, F., XU, J. and ZDEBOROVÁ, L. (2016). Mutual information in rank-one matrix estimation. *IEEE Information Theory Workshop (ITW)* 71–75.
- [28] LELARGE, M. and MIOLANE, L. Fundamental limits of symmetric low-rank matrix estimation. In *Proceedings of the 2017 Conference on Learning Theory*. PMLR 65.
- [29] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.* **7** 073403. [MR3683819](#)
- [30] MÉZARD, M., PARISI, G. and VIRASORO, M. A. (1987). *Spin Glass Theory and Beyond. World Scientific Lecture Notes in Physics* **9**. World Scientific, Teaneck, NJ. [MR1026102](#)
- [31] MIOLANE, L. (2017). Fundamental limits of symmetric low-rank matrix estimation: The non-symmetric case. ArXiv e-prints.
- [32] MONTANARI, A., REICHMAN, D. and ZEITOUNI, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inform. Theory* **63** 1572–1579. [MR3625981](#)
- [33] MONTANARI, A. and RICHARD, E. (2014). A statistical model for tensor pca. In *Neural Information Processing Systems*. 2897–2905.
- [34] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inform. Theory* **62** 1458–1484. [MR3472260](#)
- [35] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. [MR3113808](#)
- [36] PANCHENKO, D. (2013). The Parisi ultrametricity conjecture. *Ann. of Math. (2)* **177** 383–393. [MR2999044](#)
- [37] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick Model. Springer Monographs in Mathematics*. Springer, New York. [MR3052333](#)
- [38] PANCHENKO, D. (2014). The Parisi formula for mixed  $p$ -spin models. *Ann. Probab.* **42** 946–958. [MR3189062](#)
- [39] PANCHENKO, D. (2015). The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.* **43** 3494–3513. [MR3433586](#)
- [40] PANCHENKO, D. (2018). Free energy in the mixed  $p$ -spin models with vector spins. *Ann. Probab.* **46** 865–896. [MR3773376](#)
- [41] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. [MR2221787](#)
- [42] PERRY, A., WEIN, A. S. and BANDEIRA, A. (2017). Statistical limits of spiked tensor models. ArXiv e-prints.
- [43] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. [MR3845022](#)
- [44] REEVES, G. and PFISTER, H. D. (2016). The replica-symmetric prediction for compressed sensing with gaussian matrices is exact. ArXiv e-prints.
- [45] TALAGRAND, M. (2006). The Parisi formula. *Ann. of Math. (2)* **163** 221–263. [MR2195134](#)
- [46] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses. Volume I. Basic Examples*. Springer, Berlin. [MR2731561](#)

- [47] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses. Volume II. Advanced Replica-Symmetry and Low Temperature*. Springer, Heidelberg. [MR3024566](#)

SCHOOL OF MATHEMATICS  
UNIVERSITY OF MINNESOTA  
206 CHURCH ST. SE  
MINNEAPOLIS, MINNESOTA 55455  
USA  
E-MAIL: [wkchen@umn.edu](mailto:wkchen@umn.edu)