

University of Groningen

## Continuous spin mean-field models

Kulske, Christof; Opoku, Alex A.

*Published in:*  
Journal of Mathematical Physics

*DOI:*  
[10.1063/1.3021285](https://doi.org/10.1063/1.3021285)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2008

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Kulske, C., & Opoku, A. A. (2008). Continuous spin mean-field models: Limiting kernels and Gibbs properties of local transforms. *Journal of Mathematical Physics*, 49(12), [125215].  
<https://doi.org/10.1063/1.3021285>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Continuous spin mean-field models: Limiting kernels and Gibbs properties of local transforms

Christof Külske<sup>a)</sup> and Alex A. Opoku<sup>b)</sup>

*Institute of Mathematics and Computing Science, University of Groningen,  
Postbus 407, 9700 AK Groningen, The Netherlands*

(Received 3 June 2008; accepted 15 October 2008; published online 22 December 2008)

We extend the notion of Gibbsianness for mean-field systems to the setup of general (possibly continuous) local state spaces. We investigate the Gibbs properties of systems arising from an initial mean-field Gibbs measure by application of given local transition kernels. This generalizes previous case studies made for spins taking finitely many values to the first step in the direction to a general theory containing the following parts: (1) A formula for the limiting conditional probability distributions of the transformed system (it holds both in the Gibbs and in the non-Gibbs regime and invokes a minimization problem for a “constrained rate function”), (2) a criterion for Gibbsianness of the transformed system for initial Lipschitz–Hamiltonians involving concentration properties of the transition kernels, and (3) a continuity estimate for the single-site conditional distributions of the transformed system. While (2) and (3) have provable lattice counterparts, the characterization of (1) is stronger in mean field. As applications we show short-time Gibbsianness of rotator mean-field models on the  $(q-1)$ -dimensional sphere under diffusive time evolution and the preservation of Gibbsianness under local coarse graining of the initial local spin space. © 2008 American Institute of Physics. [DOI: 10.1063/1.3021285]

### I. INTRODUCTION

The study of the (failure of the) Gibbs property is a source of interesting probability theory and is linked to the study of phase transitions. Gibbs measures and generalized Gibbs measures are of interest not only on the lattice but also on more general structures. Examples of such structures are random graphs or, in the simplest conceivable case, the complete graph, where the models are called mean-field models.<sup>8,15,16,21,22,27,28</sup>

The Gibbs property of a given measure should be viewed as a continuity property of conditional probabilities as a function of the conditioning. When one tries to prove or disprove this property for a measure obtained by an application of a deterministic or stochastic transformation from a well-understood initial measure, one is led to a constrained (or “quenched”) problem, with “quenched impurities” that are induced by the conditioning. This introduces a “random” (or in better words constrained) system that we need to understand,<sup>10</sup> and this ties the problem to disordered systems and statistical mechanics on random structures.

It was through rigorous implementation of renormalization group transformations that it was discovered that images of Gibbs measures can be non-Gibbs.<sup>10,18,19,14</sup> After this discovery, there has been interest in recent times, in particular, in the study of the loss and possible recovery of the Gibbs property of an initial Gibbs measure under a stochastic time evolution. The study started in Ref. 9 where the authors focused on the evolution of an initial Gibbs measure of a lattice spin Ising model under high-temperature spin-flip Glauber dynamics. The main phenomenon observed

<sup>a)</sup>Electronic mail: kuelske@math.rug.nl. URL: <http://www.math.rug.nl/~kuelske/>.

<sup>b)</sup>Electronic mail: a.opoku@math.rug.nl.

here was the loss of the Gibbs property after a certain transition time when the system was started from an initial low-temperature state. The measure stays non-Gibbs forever when the initial external field is zero. More complicated transitions between Gibbs and non-Gibbs are possible at intermediate times when there is no spin-flip symmetry. The case of sitewise independent diffusions of continuous spins on the lattice starting from the Gibbs measure of a special double-well potential was considered in Ref. 26, exhibiting similarities and differences to the Ising case. In Ref. 11 the authors studied models for continuous compact spins, namely, the planar rotor models on the circle subjected to diffusive time evolution. It is shown therein that the time-evolved measure for high- or infinite- temperature dynamics starting from an initial arbitrary-temperature Gibbs measure will stay Gibbs for small times. The Gibbs property of the time-evolved measure is also showed to be preserved at all times for high- or infinite-temperature dynamics and initial high- or infinite-temperature Gibbs measure. The authors further proved that the time-evolved measure fails to be Gibbsian after some time when the dynamics is infinite temperature starting from an initial low-temperature Gibbs measure. Their analysis uses the machinery of cluster expansions, as earlier developed in Ref. 4. Even before it was shown that the whole process of space-time histories can be viewed as a Gibbs measure<sup>5</sup> which, however, does not imply that fixed-time projections are Gibbs.

Let us move from concrete examples to the elements of a general theory which have been proven so far. In Ref. 24 the preservation of the Gibbs property for compact (discrete and continuous) spin models for general initial interactions (having a finite “triple norm”) subjected to general sitewise transformations is studied. The technique employed therein is Dobrushin uniqueness,<sup>6,17</sup> which is quite robust and gives rise to explicit estimates. We obtained both quantitative estimates on the parameter regimes where Gibbsianness provably holds and, as the main new part, explicit continuity estimates for the conditional probabilities of the transformed system.

As an example it is shown therein that starting with an initial Gibbs measure of a rotator spin model on the  $(q-1)$ -dimensional sphere ( $q \geq 2$ ) and performing sitewise independent diffusive time evolutions, the Gibbs property is preserved in an explicitly computable time interval starting from zero. Similar conclusions were drawn for Gibbs measures for general initial interactions (with compact metric local spin spaces) subjected to a local coarse-graining transformation. (Given a decomposition of the local state space  $S$  into countably many sets, the corresponding local coarse graining is the map that associates with any point in  $S$  the label of the corresponding set in the decomposition.) Here the Gibbs property is preserved whenever the diameter of the largest set in the decomposition is small enough. Roughly speaking, this result can be seen as stability of Gibbsianness under application of a ball of sufficiently fine local transformations of coarse-graining type.

In a related line of research, transforms of initial Gibbs measures for various mean-field models were investigated. A variety of measures has been found to be non-Gibbs<sup>25,23,20</sup> in the mean-field sense. Usually the analysis of such systems shows parallels to what can be done on the lattice but goes much further. We remark that in all the cases studied so far, mean-field spins that take finitely many values had been considered, and a unifying treatment including discrete and continuous spins had been lacking. For state of the art reviews on Gibbsianness and non-Gibbsianness we refer the reader to Refs. 10 and 12.

Now, in this note we present a systematic investigation of the Gibbs property of mean-field measures subjected to local kernels. We are out to extend previous results on spins taking finitely many values to general possibly continuous (but compact) spins. More mathematical care is needed since we consider distributions of empirical measures taking values in an infinite-dimensional space. So, let us provide an informal road map of the present paper now, leaving the precise definitions and statements of the theorems to the main body.

What are the initial measures we are dealing with? We start in Sec. II by defining a class of interactions  $\Phi(\nu)$  as functions on empirical measures  $\nu$  of the system. The corresponding mean-field Hamiltonian in a volume of size  $N$  is  $N\Phi(\nu)$ . The densities of the finite-volume Gibbs measures with respect to an *a priori* product measure  $\alpha$  in volume  $N$  are given in terms of the normalized exponential  $(1/Z_N)\exp(-N\Phi(\nu))$ .

The first decision to be made is to find an appropriate notion of regularity of allowed interactions  $\Phi$ . It turns out here that the natural requirement (the suitable mean-field analog of the standard notion of absolute summability for interactions on the lattice) is that of continuous differentiability (in the space of measures on the single-site configuration space).

Next, we define the notion of mean-field Gibbsianness of a model which is given in terms of the sequence of its finite-volume measures by looking at large-volume limits of single-site conditional distributions obtained from this. This procedure provides us with a kernel  $\gamma_1(d\sigma_i|\nu)$ , where  $\nu$  is the empirical measure of a configuration in the conditioning. The model is called Gibbs if every  $\nu$  is a continuity point of  $\gamma_1$  (for the weak topology). This is a natural generalization from the discrete spin examples where this notion had been introduced and investigated before. From this definition it is also apparent that the regularity requirement on the interaction made above is natural since it implies Gibbsianness of the initial system [see more on this below (Definition 2.3)].

The situation is easier (and thus amenable to analysis) for mean-field models than for lattice models, since a configuration in the conditioning is replaced by a measure  $\nu$  on the local spin space, and it is just one single-site kernel that captures the large-volume behavior.

In Sec. III we turn to the main focus of the paper, namely, two-layer models, obtained by applying a deterministic or stochastic kernel independently over the sites to the initial model. The transformations we consider include infinite-temperature dynamics of mean-field systems with Gibbs measures as their initial distributions. In this spirit the system on the second layer describes the state of the system at any given time after the application of the dynamics. In the language of renormalization group transformations, the second-layer model is a renormalized version of the first-layer model obtained via a renormalization map given by the underlying single-site kernel. A complete analysis of the Ising model in mean field under stochastic sitewise independent time evolution has been given in Ref. 25, showing the emergence of non-Gibbsianness at sharp critical times and a phenomenon called symmetry breaking in the set of bad configurations. More examples are found in Refs. 23 and 20. At first we develop the general theory which relates our desired object, the large-system limiting conditional distribution of the transformed system, to a variational problem. In this part no specific assumptions (other than continuous differentiability of the initial potential) will be made on the model. The results hold in regions of the parameter space of the interaction where both Gibbsianness and non-Gibbsianness can occur. In the non-Gibbsian regime, however, we have to stay away from the specific critical values of the conditionings for which nonunique global minimizers occur. For the convenience of the reader we briefly review some background material on large deviations we will use for our analysis. Large deviation principles (LDPs) are interesting in themselves, but from the point of view of this paper, they will just be used as a tool to treat the limiting conditional probabilities. The main general result of this first general part is Theorem 3.10, which describes the infinite-volume second-layer conditional probabilities in terms of a solution of a variational problem (leading to a consistency equation) for the constrained first-layer model (CFLM).

In Sec. IV we provide criteria for Gibbsianness of the transformed model. This part is based on the study of the constrained consistency equation obtained in the first part of the paper. By the Tychonov theorem there exists at least one solution. By the contraction mapping theorem there is precisely one solution, provided the respective kernel is Lipschitz uniformly in the conditioning, with a constant  $L$  which can be derived explicitly when  $L < 1$ . Uniqueness of the solution implies mean-field Gibbsianness of the transformed model by the first part. This is in nice analogy to the corresponding lattice results obtained in the paper in Ref. 24 using techniques based on the Dobrushin uniqueness. More can be said, however, about the transformed system and can be put in perspective with the corresponding lattice results.

In Ref. 24 we were proving Gibbsianness but we did more than that. We provided explicit continuity estimates of the form

$$\|\gamma'_i(d\eta_i|\eta_{i^c}) - \gamma'_i(d\eta_i|\eta'_{i^c})\| \leq \sum_{j:j \neq i} Q_{i,j} d'(\eta_j, \eta'_j),$$

where  $\gamma'_i(d\eta_i|\eta_{i^c})$  are single-site conditional probabilities for the transformed system,  $Q_{ij}$  is the so-called goodness matrix, and  $d'$  is the so-called posterior metric. The posterior metric is the variational distance between constrained single-site measures  $d'(\eta_i, \eta'_i) = \|K(d\sigma_i|\eta_i) - K(d\sigma_i|\eta'_i)\|$ , where  $K$  is the joint single-site *a priori* measure (obtained in terms of  $\alpha$ , the single-site *a priori* measure for the first-layer model, and the transformation kernel). The goodness matrix  $Q_{i,j}$  is a non-negative matrix with  $\sup_{i \in G} \sum_{j \in G} Q_{i,j} < \infty$ , describing the dependence of the spin at site  $i$  on that at site  $j$ . This matrix depends on the interaction of the initial system (initial temperature) and the single-site kernel connecting the two layers. The faster the decay of  $Q_{i,j}$  is, the faster the decay of conditional probabilities on variations of the conditioning is, and the “better” or the “more Gibbsian” the system of conditional probabilities is.

In the present mean-field setup we prove as the main result of the second part of the paper an estimate of the form

$$\|\gamma'_i(d\eta_i|\nu) - \gamma'_i(d\eta_i|\nu')\| \leq L_2 \|\nu - \nu'\|,$$

with  $L_2$  given in Theorem 4.3. In the lattice estimate there is a matrix  $Q$  appearing, describing the spatial decay of influence of a variation of the conditioning at site  $j$ , while in the mean-field estimate we are simply considering the variational distance of the empirical measure of the conditioning.

$L_2$  will be finite for an initial interaction that is arbitrarily large but Lipschitz when the constrained single-site measures have good concentration properties. This is the case, e.g., at short times for diffusive time evolutions or for sufficiently fine local coarse grainings. When the initial interaction is small the transformation plays no role, and  $L_2$  is finite always.

We conclude the paper with the discussion of stochastic time evolutions and local coarse grainings in Sec. V.

## II. GENERALITIES ON MEAN-FIELD MODELS

### A. Setup

Let  $(S, d)$  and  $(S', d')$  be two given compact Polish spaces (compact separable metric spaces), each equipped with their corresponding Borel  $\sigma$ -algebras. We denote by  $\mathcal{P}(S)$ ,  $\mathcal{M}_+(S)$ , and  $\mathcal{M}(S)$  [ $\mathcal{P}(S')$ ,  $\mathcal{M}_+(S')$ , and  $\mathcal{M}(S')$ ] the spaces of probability measures, finite positive measures, and finite signed measures on  $S$  ( $S'$ ), respectively. Let  $\alpha$  and  $\alpha'$  be two given reference Borel probability measures (also called the *a priori measures*) on  $S$  and  $S'$ , respectively. In the following we will refer to  $S$  as the *initial (first-layer) single-site spin space* and  $S'$  as the *transformed (second-layer) single-site spin space*. We write  $\Omega = S^{\mathbb{N}}$  and  $\Omega' = S'^{\mathbb{N}}$ , respectively, as the configuration spaces for the initial (first layer) and the transformed (second layer) systems. In the sequel we will write probability measures for the transformed system with primes and those for the joint system (comprising of the initial and transformed systems) with tildes. The probability measures for the initial system will always be written without primes and tildes. Again we denote by  $\sigma$ ,  $\eta$ , and  $\xi$  the spin variables for the initial, the transformed, and the joint systems, respectively [e.g.,  $\xi = (\sigma_i, \eta_i)_{i \in \mathbb{N}} \in \tilde{\Omega} = (S \times S')^{\mathbb{N}}$ ]. We further set  $V_N = \{1, \dots, N\}$  and write  $\sigma_{V_N}$  for points in the product space  $S^N$ . We will simply write  $\sigma$  instead of  $\sigma_N$ . Whenever  $n < N$ , we will with abuse of notation write  $V_{N-n}$  for the set  $\{n+1, n+2, \dots, N\}$ . We now define the following concept of mean-field interaction for the initial systems that we shall consider in this work.

**Definition 2.1:** We shall refer to a map  $\Phi: \mathcal{M}_+(S) \rightarrow \mathbb{R}$  as a *proper mean-field interaction* if it satisfies the following conditions:

- (1) it is weakly continuous,
- (2) it satisfies the uniform directional differentiability condition, meaning that, for each  $\nu \in \mathcal{M}_+(S)$ , the derivative  $\Phi^{(1)}(\nu, \mu)$  at  $\nu$  in direction  $\mu$  exists and we have

$$\Phi(\nu + \mu) - \Phi(\nu) - \Phi^{(1)}(\nu, \mu) = r(\mu), \quad (1)$$

with  $\lim_{t \rightarrow 0^+} (r(t\mu)/t) = 0$  uniformly in  $\mu \in \mathcal{M}(S)$  for which  $\nu + t\mu \in \mathcal{M}_+(S)$  for  $t \in (0, 1]$ , and (3)  $\Phi^{(1)}(\nu, \mu)$  is a continuous function of  $\nu$ .

For each mean-field interaction  $\Phi$  and each  $N \in \mathbb{N}$  we define the finite-volume Hamiltonian  $H_N$  (a real-valued function on the product space  $S^N$ ) as

$$H_N(\sigma_{V_N}) := N\Phi(L_N(\sigma_{V_N})), \quad (2)$$

where  $L_N(\sigma_{V_N}) = (1/N) \sum_{i=1}^N \delta_{\sigma_i}$  is the empirical measure. Observe from the permutation invariance of the empirical measures that  $H_N$  is also permutation invariant. With this notation we define for each  $N \in \mathbb{N}$  the *finite-volume mean-field model*  $\mu_{\beta, N}$  associated with the finite-volume Hamiltonian  $H_N$  and an inverse temperature  $\beta$  as

$$\mu_{\beta, N}(d\sigma_{V_N}) := \frac{e^{-\beta H_N(\sigma_{V_N})} \alpha^{\otimes N}(d\sigma_{V_N})}{\int_{S^N} e^{-\beta H_N(\sigma_{V_N})} \alpha^{\otimes N}(d\bar{\sigma}_{V_N})}. \quad (3)$$

Here we have used  $\otimes$  to denote tensor product of measures. In the following, unless otherwise stated, the inverse temperature  $\beta$  will be absorbed into the interaction  $\Phi$ . In view of this, we will write  $\mu_N$  instead of  $\mu_{\beta, N}$ . It follows from the permutation invariance of the  $H_N$  that the measure  $\mu_N$  is invariant under permutation of its arguments. We call the sequence of permutation invariant probability measures  $(\mu_N)_{N \in \mathbb{N}}$  mean-field model on  $S$  associated with  $\Phi$  and  $\alpha$ . Furthermore, with abuse of notation we will in the sequel write  $\mu_N$  for  $(\mu_N)_{N \in \mathbb{N}}$ . This sequence according to the de Finetti theorem has weak limit  $\mu$  which is a convex combination of product measures.<sup>13</sup> A variational characterization of these infinite-volume measures and related results will be the content of another paper which will appear elsewhere.

## B. Notion of Gibbsianness for mean-field models

In this subsection we aim at prescribing a general notion of Gibbsianness for mean-field models, extending a similar notion given earlier in Refs. 23 and 25 for the corresponding Ising spins to more general (possibly continuous) spins. Let us start by recalling the notion of Gibbsianness for lattice spin models. An infinite-volume lattice probability measure  $\mu$  is said to be Gibbs if its finite-volume conditional distributions are non-null uniformly with respect to the conditioning or the boundary condition (i.e., for each boundary condition the conditional distributions assign nonzero measures to any nonempty open set) and continuous with respect to the topology generated by local observables as a function of the conditioning. It is important to note that the conditionings are configurations living on the complements of finite volumes, i.e., they are infinite-volume configurations. In practice one starts with a certain family of (candidate) conditional distributions called a specification, which satisfies the above conditions and takes limits along appropriate boundary conditions to obtain the desired infinite-volume Gibbs measures  $\mu$ .

Now, how does one transfer the above notion of Gibbsianness to mean-field models? Well, as in the above, we need to consider finite-volume conditional distributions with infinite-volume conditionings (depicting the influence of the rest of the system on the chosen finite volume). However, in contrast to the case on the lattice these finite-volume conditional distributions cannot be readily obtained. They are obtained via a limiting procedure starting from conditional distributions for the finite-volume mean-field model  $\mu_N$ . Next, we need to make clear how we fix boundary conditions. Observe that boundary conditions that are permutations of one another will give rise to the same effect, since the boundary conditions enter the measures of interest via the empirical measure. Due to this, it is reasonable to consider boundary conditions from the subset of the configuration space resulting from configurations which are not permutations of one another, i.e., the quotient space generated by permutation. Choosing a boundary condition from this quotient space is equivalent to fixing the empirical measure to a given probability measure on the

single-site spin space. So in the following we will always condition on the empirical measures. This means that we need a different topology, namely, the weak topology on the space of probability measures on the single-site spin space, for defining Gibbsianness for mean-field models instead of the topology of local observables considered above. In general, the infinite-volume measure for the mean-field model is a mixture of product measures. This means that all of the configurations will be points of discontinuity for the conditional distributions and hence the non-Gibbsianness of the measure under the topology on the configuration space. However, thanks to the change of topology, this subtlety is circumvented and we can now speak about Gibbsianness for mean-field models. Furthermore, this topology has the nice feature of lifting properties such as compactness and separability of the single-site spin space to the space of probability measures on the single-site spin space. In view of the above, it is reasonable to define Gibbsianness for mean-field models as follows.

**Definition 2.2:** Let  $E$  be a compact Polish space and  $\rho_N$  be a mean-field model on  $E$ . We call  $\nu \in \mathcal{P}(E)$  a good configuration for  $\rho_N$  if and only if

- (1) the limit

$$\gamma_1(dx_1|\lambda) := \lim_{N \uparrow \infty} \rho_N(dx_1|x_{V_{N-1}}), \quad (4)$$

where

$$\lambda = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{i=2}^N \delta_{x_i},$$

exists for all  $\lambda$  in a weak neighborhood of  $\nu$  and

- (2) for any Borel subset  $A \subset E$ , the function  $\lambda \mapsto \gamma_1(A|\lambda)$  is weakly continuous at  $\nu$ . We say  $\rho_N$  is Gibbs if and only if every configuration is good.

**Remark:**

- (1) For binary single-site spin space  $E$ , the measures in  $\mathcal{P}(E)$  are completely described by their mean. This makes it possible to use the empirical average or magnetization as boundary condition which is simply the same as given the empirical measure. With this simplification, the above definition reduces to the notion of Gibbsianness for the corresponding Curie–Weiss model studied in Refs. 23 and 25. There the topology of interest is the topology on the interval  $[-1, 1]$  generated by the Euclidean metric.
- (2) The object  $\gamma_1$  defines a probability kernel from  $\mathcal{P}(E)$  to  $E$ , i.e.,  $\gamma_1$  is a function from  $\mathcal{E} \times \mathcal{P}(E)$  to  $[0, 1]$ , which depends measurably on  $\mathcal{P}(E)$  for any fixed Borel measurable subset  $A \in \mathcal{E}$  of  $E$  and is a probability measure for fixed  $\nu \in \mathcal{P}(E)$ .

The above remark indicates that  $\gamma_1: \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ , and hence we can talk about its fixed points. These fixed points necessitate the following notion of consistency for mean-field models.

**Definition 2.3:** A probability measure  $\nu \in \mathcal{P}(E)$  is said to be consistent with respect to  $\gamma_1$  whenever

$$\gamma_1(dx_1|\nu) = \nu(dx_1). \quad (5)$$

In what follows (unless otherwise stated) all topological considerations on the space of probability measures will be with respect to the weak topology. We now formulate a result regarding the Gibbsianness of the mean-field model  $\mu_N$  on  $S$  associated with a mean-field interaction  $\Phi$  and  $\alpha$ .

**Proposition 2.4:** Let  $\mu_N$  be as above. Then

- (1) for each  $n \in \mathbb{N}$  and each  $\nu \in \mathcal{P}(S)$ , the finite-volume conditional distribution  $\gamma_n(\cdot|\nu)$  for the mean-field model  $\mu_N$  with  $\nu$  as infinite-volume boundary condition is given by

$$\gamma_n(d\sigma_{V_n}|\nu) := \lim_{N \uparrow \infty} \mu_N(d\sigma_{V_n}|\sigma_{V_{N-n}}) = \gamma_1^{\otimes n}(d\sigma_1|\nu), \tag{6}$$

where

$$\nu = \lim_{N \uparrow \infty} \sum_{i=n+1}^N \delta_{\sigma_i} \quad \text{and} \quad \gamma_1(d\sigma_1|\nu) = \frac{e^{-\Phi^{(1)}(\nu, \delta_{\sigma_1})} \alpha(d\sigma_1)}{\int_S e^{-\Phi^{(1)}(\nu, \delta_{\bar{\sigma}_1})} \alpha(d\bar{\sigma}_1)},$$

- (2) the mean-field model  $\mu_N$  is Gibbs, and
- (3)  $\gamma_1$  admits at least one consistent measure.

**Remark:** It is important to note that in the study of Gibbsianness for mean-field models we need only one quantity, namely,  $\gamma_1$ , as opposed to the corresponding case on the lattice where one has to investigate the continuity properties of all the single-site conditional distributions.

**Proof:**

- (1) Let us start with the proof of the first assertion. Let us choose  $n, N \in \mathbb{N}$  such that  $N > n$ . Then we can write the empirical measure  $L_N(\sigma_{V_N})$  as

$$L_N(\sigma_{V_N}) = \frac{n}{N} L_n(\sigma_{V_n}) + \frac{N-n}{N} L_{N-n}(\sigma_{V_{N-n}}), \tag{7}$$

where  $L_{N-n}(\sigma_{V_{N-n}}) = [1/(N-n)] \sum_{i=n+1}^N \delta_{\sigma_i}$ . Observe from (3) that the conditional distribution  $\mu_N(\cdot | \sigma_{V_{N-n}})$  is unchanged if we add  $-N\Phi([ (N-n)/N ] L_{N-n}(\sigma_{V_{N-n}}))$  to the Hamiltonian  $H_N$  defining  $\mu_N$ . Thus we can use in the definition of  $\mu_N(\cdot | \sigma_{V_{N-n}})$  the Hamiltonian

$$\begin{aligned} H_{n,N}(\sigma_{V_N}) &:= N \left\{ \Phi \left( \frac{N-n}{N} L_{N-n}(\sigma_{V_{N-n}}) + \frac{n}{N} L_n(\sigma_{V_n}) \right) - \Phi \left( \frac{N-n}{N} L_{N-n}(\sigma_{V_{N-n}}) \right) \right\} \\ &= N \left\{ \Phi^{(1)} \left( \frac{N-n}{N} L_{N-n}(\sigma_{V_{N-n}}), \frac{n}{N} L_n(\sigma_{V_n}) \right) - r \left( \frac{n}{N} L_n(\sigma_{V_n}) \right) \right\} \\ &= \sum_{i=1}^n \Phi^{(1)} \left( \frac{N-n}{N} L_{N-n}(\sigma_{V_{N-n}}), \delta_{\sigma_i} \right) - Nr \left( \frac{n}{N} L_n(\sigma_{V_n}) \right). \end{aligned} \tag{8}$$

In the second equality above we have made use of property (1) of  $\Phi$  and in the third we employed the linearity of  $\Phi^{(1)}$  in its second argument and the fact that  $L_n(\sigma_{V_n}) = (1/n) \sum_{i=1}^n \delta_{\sigma_i}$ . It is observed from (2) of Definition 2.1 that  $\lim_{N \uparrow \infty} Nr((n/N)L_n(\sigma_{V_n})) = 0$  since we have set  $t = 1/N$ . Therefore by replacing  $H_N$  in (3) by  $H_{n,N}$  and exploiting the continuity property of  $\Phi^{(1)}$  in its first argument, we arrive at the desired expressions for  $\gamma_n$  and  $\gamma_1$  after taking an  $N \rightarrow \infty$  limit and conditioning on  $\nu = \lim_{N \uparrow \infty} \sum_{i=n+1}^N \delta_{\sigma_i}$ .

- (2) The existence of  $\gamma_1(\cdot | \lambda)$  for any  $\lambda \in \mathcal{P}(S)$  follows from the proof of assertion (1). The continuity property of the kernel  $\gamma_1(\cdot | \cdot)$  in its second argument also follows trivially from the continuity property of  $\Phi^{(1)}$  with respect to its first argument, hence the proof that the mean-field model  $\mu_N$  is Gibbs.
- (3) The existence of a fixed point for  $\gamma_1$  follows from Tychonov’s fixed point theorem, which states that for any nonempty compact convex subset  $X$  of a locally convex topological vector space  $V$  and continuous function  $f: X \rightarrow X$ , there is a fixed point for  $f$ .

Now note that  $\gamma_1: \mathcal{P}(S) \rightarrow \mathcal{P}(S)$  and that  $\mathcal{P}(S)$  is a compact convex subset of  $\mathcal{M}(S)$ , the space of finite signed measures on  $S$ , since  $S$  is compact and by the weak topology so is  $\mathcal{P}(S)$ . Therefore we only need to show that  $\mathcal{M}(S)$  is a locally convex topological vector space with respect to the weak topology on  $\mathcal{M}(S)$ . We show this via the following sequence of arguments: The total variational norm turns the space  $\mathcal{M}(S)$  into a Banach space. This then implies that  $\mathcal{M}(S)$  is normable and consequently it is locally convex (i.e., the origin has a local base of convex sets)

with respect to the total variational topology on  $\mathcal{M}(S)$  by Theorem 1.39 of Ref. 29. It also follows from a corollary to Theorem 3.4 of Ref. 29 that the dual space  $\mathcal{M}^*(S)$  of  $\mathcal{M}(S)$  separates points in  $\mathcal{M}(S)$ . Finally, by Theorem 3.10 of Ref. 29 the weak topology on  $\mathcal{M}(S)$  generated by the dual space  $\mathcal{M}^*(S)$  turns  $\mathcal{M}(S)$  into a locally convex topological space.  $\square$

On the lattice it is known that any translation invariant uniformly absolutely summable interaction gives rise to a quasilocal specification, but there are subtleties in the converse: A quasilocal specification can be written with the aid of an absolutely summable potential only if one gives up translation invariance. Similarly, in our mean-field case, it can be asked for the precise assumption on the  $\gamma$  such that a nicely behaved interaction can be constructed. We postpone this question to a later investigation.

The aim for the rest of the paper is to study the Gibbs properties of transforms of initial Gibbsian mean-field models. This is based on studying the continuity property of the single-site kernel  $\gamma'_1$  for the transformed system. In our investigation of this continuity property for  $\gamma'_1$  we will employ the machinery of large deviations theory. In view of this, we recall some basic facts about large deviations theory, that we shall use in our study, in Sec. II C.

### C. Some facts about large deviations theory

In this subsection we recall some facts about large deviations theory, and for a detailed discussion on this theory and its application to statistical mechanics we refer the reader to Refs. 3 and 7. Let  $X$  be a Polish space equipped with its Borel  $\sigma$ -algebra.

**Definition 2.5:** A sequence of probability measures  $(Q_N)_{N \in \mathbb{N}}$  in  $\mathcal{P}(X)$  is said to satisfy a LDP on  $X$  with rate  $a_N$  (sequence of positive numbers tending to infinity) and rate function  $I: X \rightarrow [0, +\infty]$  if

- (1)  $I$  is lower semicontinuous on  $X$  and the level sets  $\{x \in X: I(x) \leq a\}$  are compact for all  $a \in [0, +\infty)$ ;
- (2) for any Borel subset  $B$  of  $X$ ,

$$-I(\overset{\circ}{B}) \leq \liminf_{N \rightarrow \infty} a_N^{-1} \log Q_N(B) \leq \limsup_{N \rightarrow \infty} a_N^{-1} \log Q_N(B) \leq -I(\bar{B}), \quad (9)$$

where for any subset  $C$  of  $X$ ,  $I(C) = \inf_{x \in C} I(x)$ , and  $\overset{\circ}{C}$  and  $\bar{C}$  are, respectively, the interior and the closure of  $C$ .

As an example take  $(Y_n)_{n \in \mathbb{N}}$ , an independent and identically distributed sequence of random variables on  $X$  with  $\rho$  as the law of  $Y_1$ . Let  $Q_N$  be the distribution of the empirical measures  $L_N = (1/N) \sum_{i=1}^N \delta_{Y_i}$ . Then  $Q_N$  satisfies LDP with rate  $N$  and rate function

$$S(\nu|\rho) = \begin{cases} \int \frac{d\nu}{d\rho} \log \frac{d\nu}{d\rho} d\rho & \text{if } \nu \ll \rho \text{ and } \frac{d\nu}{d\rho} \log \frac{d\nu}{d\rho} \in \mathcal{L}^1(\rho) \\ +\infty & \text{otherwise,} \end{cases} \quad (10)$$

where  $S(\cdot|\cdot)$  is the relative entropy and  $d\nu/d\rho$  is the Radon–Nikodym derivative of  $\nu$  given  $\rho$ . The above example is Sanov's theorem in large deviations theory as can be found, e.g., in Theorem II.4.3 of Ref. 7.

Another important fact about LDP that we shall employ in our study is the *contraction principle* (see, e.g., Theorem II.5.1 of Ref. 7), which comes into play when one is concerned with partial summary of the information weighted by  $Q_N$ . More precisely, suppose  $\psi$  is a continuous function from the Polish space  $X$  to another Polish space  $Y$  and  $Q_N$  is a sequence of probability measures on  $X$  satisfying the LDP with rate  $a_N$  rate function  $I$ . Then the sequence  $\hat{Q}_N = Q_N \circ \psi^{-1}$  of probability measures on  $Y$  also satisfies LDP with rate  $a_N$  and rate function  $\hat{I}$  given by

$$\hat{I}(y) = \inf\{I(x): \psi(x) = y\}. \quad (11)$$

Our last fact from LDP concerns the integrals of exponentials of functionals of random variables whose distributions satisfy LDP. This is found, e.g., in Ref. 7 as Theorem II.7.2a. The result in Ref. 7 is more general than what is stated here.

**Fact 2.6:** *Let  $X$  be a Polish space and  $Q_N$  a sequence of probability measures on  $X$  obeying LDP with rate  $a_N$  and rate function  $I$ . Suppose that  $F: X \rightarrow \mathbb{R}$ , which is continuous and bounded below, and for each  $N \in \mathbb{N}$  the integral  $\int_X \exp(-a_N F(x)) Q_N(dx)$  is finite. Let  $Q_{N,F}$  be the sequence of probability measures given by*

$$Q_{N,F}(A) = \frac{\int_A \exp(-a_N F(x)) Q_N(dx)}{\int_X \exp(-a_N F(x)) Q_N(dx)} \tag{12}$$

for any Borel subset  $A$  of  $X$ . Then  $Q_{N,F}$  satisfies LDP with rate  $a_N$  and rate function

$$I_F(x) = I(x) + F(x) - \inf_{y \in X} (I(y) + F(y)). \tag{13}$$

### III. TWO-LAYER SYSTEM AND GIBBSIANNES OF TRANSFORMED SYSTEMS

We now introduce on  $S \times S'$  a Borel probability measure  $K$  such that

$$K(d\xi_i) = k(\sigma_i, \eta_i) \alpha(d\sigma_i) \alpha'(d\eta_i), \tag{14}$$

with

$$\sup_{(\sigma_i, \eta_i) \in S \times S'} |\log k(\sigma_i, \eta_i)| < \infty.$$

We assume further that  $\alpha = \int k(\cdot, \eta_i) \alpha'(d\eta_i)$  and  $\alpha' = \int k(\sigma_i, \cdot) \alpha(d\sigma_i)$ , where we are using the subscript  $i \in \mathbb{N}$  to convey the idea that  $K$  is the joint *a priori* measure for site  $i$ . As discussed in Ref. 24 we begin with a first-layer mean-field system, described by some given mean-field interaction  $\Phi$  and the *a priori* measure  $\alpha$  with  $S$  as its single-site space. This system is coupled to a second system (second-layer system with  $S'$  as its single-site spin space) via  $K$ . In other words, we begin with two independent systems, namely, the first-layer system, described by  $\Phi$  and  $\alpha$ , and the second-layer system which is independent and identically distributed with distribution  $\alpha'$ . Note that the components of the two systems are indexed by  $\mathbb{N}$ , the set of positive integers. These two systems are then coupled vertically via the kernel  $k$  (14), giving rise to our so-called two-layer (joint) system. Then as before (3) the mean-field model  $\tilde{\mu}_N$  for our two-layer system in a finite volume  $V_N$  is given by

$$\begin{aligned} \tilde{\mu}_N(d\xi_{V_N}) &= \frac{\exp(-N\Phi(L_N(\sigma_{V_N}))) \prod_{i=1}^N K(d\sigma_i, d\eta_i)}{\int_{(S \otimes S')^N} \exp(-N\Phi(L_N(\hat{\sigma}_{V_N}))) \prod_{i=1}^N K(d\hat{\sigma}_i, d\hat{\eta}_i)} \\ &= \frac{\exp(-N\{\Phi(\pi_1 L_N(\xi_{V_N})) - L_N(\xi_{V_N})[\log k(\cdot, \cdot)]\}) \prod_{i=1}^N \alpha(d\sigma_i) \alpha'(d\eta_i)}{\int_{(S \otimes S')^N} \exp(-N\{\Phi(\pi_1 L_N(\hat{\xi}_{V_N})) - L_N(\hat{\xi}_{V_N})[\log k(\cdot, \cdot)]\}) \prod_{i=1}^N \alpha(d\hat{\sigma}_i) \alpha'(d\hat{\eta}_i)}, \end{aligned} \tag{15}$$

where  $L_N(\xi_{V_N}) = (1/N) \sum_{i=1}^N \delta_{(\sigma_i, \eta_i)}$  is the joint empirical measures and  $\pi_1 L_N(\xi_{V_N}) = L_N(\sigma_{V_N})$  is the projection onto the first variable. We have denoted by  $\nu[f]$  the integral of the measurable map  $f$  with respect to the measure  $\nu$ . Furthermore, in the above we have made use of (2) and (14). Under

the continuous map  $\xi_{V_N} \mapsto L_N(\xi_{V_N})$ , the joint measures  $\tilde{\mu}_N \in \mathcal{P}((S \times S')^M)$  have unique image measures (push forwards)  $\tilde{Q}_{N,\Phi,k}$  in  $\mathcal{P}(\mathcal{P}(S \times S'))$ . So we now study the large deviation properties of the empirical measures  $L_N(\xi_{V_N})$  under  $\tilde{Q}_{N,\Phi,k}$ , and the proposition below is a summary of this LDP property.

**Proposition 3.1:** *The sequence of probability measures  $\tilde{Q}_{N,\Phi,k}$  satisfies a LDP with rate  $N$  and rate function  $\tilde{J}$  given by*

$$\tilde{J}(\tilde{\nu}) = \Phi(\pi_1 \tilde{\nu}) - \tilde{\nu}[\log k(\cdot, \cdot)] + S(\tilde{\nu} | \alpha \otimes \alpha') - \text{const}, \tag{16}$$

where

$$\text{const} = \inf_{\tilde{\lambda} \in \mathcal{P}(S \times S')} \{ \Phi(\pi_1 \tilde{\lambda}) - \tilde{\lambda}[\log k(\cdot, \cdot)] + S(\tilde{\lambda} | \alpha \otimes \alpha') \}.$$

**Remark:**

- (1) Proposition 3.1 is a direct consequence of Fact 2.6 since  $\Phi$  is bounded and the weak topology on  $\mathcal{P}(S \times S')$  turns  $\mathcal{P}(S \times S')$  into a compact separable metric space.
- (2) The continuous map  $\xi_{V_N} \mapsto L_N(\xi_{V_N})$  from  $(S \times S')^N$  to  $\mathcal{P}(S \times S')$  also induces a continuous map (say,  $\vartheta_N$ ) from  $\mathcal{P}((S \times S')^N)$  into  $\mathcal{P}(\mathcal{P}(S \times S'))$ . So the push forwards  $\tilde{Q}_{N,\Phi,k}$  are simply obtained by replacing the product measures  $\prod_{i=1}^N \alpha \otimes \alpha'$  in the second equality in (15) by their images under  $\vartheta_N$ .
- (3) Observe from Proposition 3.1 that the empirical measures of the initial spin variables under the initial finite-volume mean-field model  $\mu_N$  (3) satisfy a LDP with rate  $N$  and rate function  $I_\alpha^\Phi$  given by

$$I_\alpha^\Phi(\nu) = S(\nu | \alpha) + \Phi(\nu) - \inf_{\mu \in \mathcal{P}(S)} [S(\mu | \alpha) + \Phi(\mu)]. \tag{17}$$

**A. Transforms of Gibbsian mean-field models**

Given a Gibbs measure  $\mu$  for the initial model (3), it is our aim in the rest of this work to investigate the Gibbs properties of the transformed measures

$$\mu'(d\eta) = \int_{\Omega} \mu(d\sigma) \prod_{i \in \mathbb{N}} k(\sigma_i, \eta_i) \alpha'(d\eta_i), \tag{18}$$

as has been done for the corresponding short-range models in Ref. 24. The study, as in Ref. 24 and the above, is based on investigating the continuity properties of the single-site conditional distribution  $\gamma'_i$  for the transformed system. This consists in studying the infinite-volume  $N$ -limits of the following finite-volume quantity  $\mu'_N(d\eta_1 | \eta_{V_{N-1}})$ , where the finite-volume transformed mean-field measures  $\mu'_N$  are second marginals of the joint measures  $\tilde{\mu}_N$ , i.e.,

$$\mu'_N = \int_{S^N} \tilde{\mu}_N(d\sigma_{V_N}). \tag{19}$$

Let us now digress a bit to discuss the LDP property for the transformed measures  $\mu'_N$  which is also of interest in itself. The associated rate function consists of both entropic and energetic terms. The energetic term (or renormalized interaction) is obtained by solving some minimization problem. The unicity of the global minimizers of the underlying minimization problem will determine the Gibbsianity of the transformed system.

**1. LDP for transformed system and related results**

For any  $\tilde{\nu} \in \mathcal{P}(S \times S')$  we denote by  $\pi_2 \tilde{\nu}$  the marginal of  $\tilde{\nu}$  on  $S'$ . Denote by  $Q'_{N,\Phi,k}$  the push forward of  $\mu'_N$  given by  $Q'_{N,\Phi,k} = \tilde{Q}_{N,\Phi,k} \circ \pi_2^{-1}$ . Therefore by the contraction principle, the empirical measures of the transformed spins under  $Q'_{N,\Phi,k}$  satisfy a LDP which we formulate in the proposition below.

**Proposition 3.2:** *The sequence of probability measures  $Q'_{N,\Phi,k}$  in  $\mathcal{P}(\mathcal{P}(S'))$  satisfies a LDP with rate  $N$  and rate function  $J'$  given by*

$$J'(\nu') = \inf_{\substack{\tilde{\nu} \in \mathcal{P}(S \times S') \\ \pi_2 \tilde{\nu} = \nu'}} \tilde{J}(\tilde{\nu}) = S(\nu' | \alpha') + \inf_{\tilde{\nu} \in M_{\nu'}} J_{\nu'}(\tilde{\nu}) - \text{const}, \tag{20}$$

where

$$J_{\nu'}(\tilde{\nu}) = S(\tilde{\nu} | \alpha \otimes \nu') + \Phi(\pi_1 \tilde{\nu}) - \tilde{\nu}[\log k(\cdot, \cdot)],$$

and  $M_{\nu'}$  is the subset of  $\mathcal{P}(S \times S')$  consisting of probability measures with fixed second marginal  $\nu'$ .

**Remark:**

- (1) For each  $\nu' \in \mathcal{P}(S')$ ,  $J_{\nu'}$  up to an additive constant (depending on  $\nu'$ ) is the large deviation rate function for the joint system when the second-layer system is constrained to configurations with empirical measure  $\nu'$ , i.e.,  $J$  (up to an additive constant) is the rate function for the CFLM. This CFLM rate function will play a key role in determining whether the transformed system is Gibbs or not. We shall show below that the continuity properties of the transformed single-site kernels  $\gamma'_1(\cdot | \nu')$  will be determined by the unicity of the global minimizers of the function  $J_{\nu'}$  uniformly in  $\nu'$ . Observe also that  $J_{\nu'}$  is a lower semicontinuous function and that it also attains its infimum on  $M_{\nu'}$  since  $M_{\nu'}$  is a compact subset of  $\mathcal{P}(S \times S')$ . Additionally,  $M_{\nu'}$  is convex.
- (2) Let us set  $\Phi'(\nu') := \inf_{\tilde{\nu} \in M_{\nu'}} J_{\nu'}(\tilde{\nu})$ . Then by comparing the expression for  $J'$  with that of  $I_\alpha^\Phi$  (17), we deduce that  $\Phi'$  is the interaction for the transformed system. Thus for the transformed system to be Gibbs we do require  $\Phi'$  to satisfy the conditions in Definition 2.1.

**Proof of Proposition 3.2:** The first equality in expression (20) of Proposition 3.2 for  $J'$  follows from the contraction principle since the map  $\pi_2: \mathcal{P}(S \times S') \rightarrow \mathcal{P}(S')$  is weakly continuous. Further, for each  $\nu' \in \mathcal{P}(S')$  the set  $M_{\nu'}$  is compact because  $M_{\nu'} = \pi_2^{-1}(\{\nu'\})$ , and by continuity of  $\pi_2$ ,  $M_{\nu'}$  is a closed subset of  $\mathcal{P}(S \times S')$ . Now it also follows from standard results in analysis that closed subsets of a compact set are compact, and hence the compactness of  $M_{\nu'}$ , since by the compactness property of  $S \times S'$ , the weak topology turns  $\mathcal{P}(S \times S')$  into a compact space. Moreover,  $\tilde{J}$  is lower semicontinuous and consequently the infimum of  $\tilde{J}$  over measures with fixed second marginal  $\nu'$  is attained on  $M_{\nu'}$ .

Now for each  $\nu' \in \mathcal{P}(S')$ , the measures  $\tilde{\nu} \in M_{\nu'}$  are of the form  $\tilde{\nu}(d\xi_i) = \nu'(d\eta_i) \tilde{\nu}(d\sigma_i | \eta_i)$ , where for each  $\eta_i \in S'$ ,  $\tilde{\nu}(d\sigma_i | \eta_i) \in \mathcal{P}(S)$ . With this representation, the relative entropy for the elements in  $M_{\nu'}$  with respect to  $\alpha \otimes \alpha'$  takes the form

$$S(\tilde{\nu} | \alpha \otimes \alpha') = \int_{S \otimes S'} \nu'(d\eta_i) \tilde{\nu}(d\sigma_i | \eta_i) \log \left( \frac{d\nu'}{d\alpha'}(\eta_i) \frac{d\tilde{\nu}(\sigma_i | \eta_i)}{d\alpha} \right) = S(\nu' | \alpha') + S(\tilde{\nu} | \alpha \otimes \nu'). \tag{21}$$

The second equality of expression (20) for  $J'$  follows by putting the above form of the relative entropy into expression (16) for  $\tilde{J}$ . □

The measures in  $M_{\nu'}$  that are of interest in determining  $J'(\nu')$  are at most those for which  $\tilde{\nu}(\cdot | \eta_i) \ll \alpha$  for all  $\eta_i \in S'$ , i.e.,

$$\tilde{\nu}(d\xi_i) = \nu'(d\eta_i)\alpha(d\sigma_i)f_{\nu'}(\sigma_i|\eta_i) \tag{22}$$

for some measurable function  $f_{\nu'}: S \times S' \rightarrow [0, \infty]$  with the property that for each  $\eta_i \in S'$ ,  $\int \alpha(d\sigma_i)f_{\nu'}(\sigma_i|\eta_i) = 1$ . We call  $f_{\nu'}$  the *conditional  $\alpha$ -density* of  $\tilde{\nu}$ . This reduces the whole problem of minimizing  $J_{\nu'}$  over  $M_{\nu'}$  to the problem of finding the conditional  $\alpha$ -densities  $f_{\nu'}$  for which  $\nu'(d\eta_i)\alpha(d\sigma_i)f_{\nu'}(\sigma_i|\eta_i)$  is a minimizer of  $J_{\nu'}$ .

As our next result, we present an explicit expression for the conditional  $\alpha$ -densities  $f_{\nu'}(\sigma_i|\eta_i)$  at which  $J_{\nu'}$  attains both global and local minima.

**Theorem 3.3:** *For any  $\nu' \in \mathcal{P}(S')$ , the function  $J_{\nu'}$  attains its infimum on  $M_{\nu'}$ . Furthermore, any minimizer (global or local)  $\tilde{\nu} \in M_{\nu'}$  of  $J_{\nu'}$  has the conditional  $\alpha$ -density  $f_{\nu'}$  which is  $\alpha \otimes \nu'$  almost sure (a.s.) strictly positive and satisfies the “constrained mean-field equation”*

$$f_{\nu'}(\sigma_i|\eta_i) = \frac{e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\sigma_i})}k(\sigma_i, \eta_i)}{\int e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\hat{\sigma}_i})}k(\hat{\sigma}_i, \eta_i)\alpha(d\sigma_i)} \tag{23}$$

**Remark:**

- (1) As we pointed out in the above, the measures  $\tilde{\nu} \in M_{\nu'}$  that are involved in determining  $J'(\nu')$  are those that take the form (22). The minimizers of  $J_{\nu'}$  are among these probability measures and indeed they are those probability measures with  $f_{\nu'}$  given by (23).
- (2) Note also that the minimizers of the rate function  $I_\alpha^\Phi$  (17) for the initial system are also measures  $\nu \in \mathcal{P}(S)$  with  $\nu(d\sigma_i) = f(\sigma_i)\alpha(d\sigma_i)$ , where the  $f$  are  $\alpha$  a.s. strictly positive and satisfy the mean-field equation

$$f(\sigma_i) = \frac{e^{-\Phi^{(1)}(\nu, \delta_{\sigma_i})}}{\int e^{-\Phi^{(1)}(\nu, \delta_{\hat{\sigma}_i})}\alpha(d\hat{\sigma}_i)} \tag{24}$$

Thus minimizers of  $I_\alpha^\Phi$  are consistent measures for the kernel  $\gamma_1$  of the initial mean-field model. We defer the proof of Theorem 3.3 to the Appendix at the end of the paper.

We have seen from the second remark of the above theorem that the minimizers of the rate function  $I_\alpha^\Phi$  are consistent measures for the kernel  $\gamma_1$  of the initial mean-field model. A natural question that one would like to ask is “with respect to which kernel are the minimizers of  $J_{\nu'}$  consistent?” To answer this question, remember that the measures we are interested in are those in  $M_{\nu'}$ . Thus we need to care about consistency for only the conditional measures  $\tilde{\nu}(\cdot|\eta_i)$ . This leads to the following choice of probability kernel from  $M_{\nu'}$  to  $S \times S'$ .

**Definition 3.4:** *We refer to the map  $\gamma_{\nu'}: M_{\nu'} \rightarrow M_{\nu'}$  given by*

$$\gamma_{\nu'}(d\xi_i|\tilde{\nu}) = \nu'(d\eta_i) \frac{e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\sigma_i})}k(\sigma_i, \eta_i)\alpha(d\sigma_i)}{\int e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\hat{\sigma}_i})}k(\hat{\sigma}_i, \eta_i)\alpha(d\sigma_i)} \tag{25}$$

as the *constrained first-layer probability kernel (CFLPK)*.

**Proposition 3.5:** *For any probability measure  $\nu' \in \mathcal{P}(S')$ , the CFLPK  $\gamma_{\nu'}$  has a fixed point.*

**Remark:** Observe that not all the measures  $\tilde{\nu} \in M_{\nu'}$  have  $\tilde{\nu}(\cdot|\cdot) \ll \alpha$ , but the fixed points of  $\gamma_{\nu'}$  do and include all the minimizers of  $J_{\nu'}$ .

Thus the consistent probability measures for the CFLPK are those measures  $\tilde{\nu} \in M_{\nu'}$  for which  $\tilde{\nu}(\cdot|\cdot) \ll \alpha$  and have conditional  $\alpha$ -density functions  $f_{\nu'}$  (23).

**Proof:** Note that for any Borel subset  $A$  of  $S \times S'$ , the map  $\gamma_{\nu'}(A|\cdot): M_{\nu'} \rightarrow \mathbb{R}$  is continuous by the continuity property of  $\Phi^{(1)}$ .

The existence of a fixed point for the CFLPK follows from Tychonov’s fixed point theorem as explained in the proof of assertion (3) of Proposition 2.4 since we know from the proof of

Proposition 3.2 that for any  $\nu' \in \mathcal{P}(S')$ , the set  $M_{\nu'}$  is compact with respect to the weak topology. It is also not hard to see that  $M_{\nu'}$  is convex. Finally, it follows from the arguments employed in the proof of assertion (3) of Proposition 2.4 that  $\mathcal{M}(S \times S')$ , the set of finite signed measures on  $S \times S'$ , is locally convex topological vector space under the weak topology.  $\square$

Note also that for each  $\nu' \in \mathcal{P}(S')$ , the CFLM is Gibbsian (in the sense of Definition 2.2) because of the continuity property of  $\gamma_{\nu'}$ . For each  $\nu' \in \mathcal{P}(S')$  denote by  $\mathcal{C}_{\nu'}$  the set of all consistent probability measures of  $\gamma_{\nu'}$ . We state as our next result the following lemma concerning a single-site variational principle for the CFLM.

**Lemma 3.6:** *For any probability measure  $\tilde{\nu} \in M_{\nu'}$ , the relative entropy  $S(\tilde{\nu}|\alpha \otimes \nu')$  satisfies*

$$S(\tilde{\nu}|\alpha \otimes \nu') \geq -\Phi^{(1)}(\pi_1 \tilde{\nu}, \pi_1 \tilde{\nu}) + \tilde{\nu}[\log k(\cdot, \cdot)] - \int \nu'(d\eta_i) \log \int \alpha(d\sigma_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\sigma_i})} k(\delta_{\sigma_i}, \eta_i). \tag{26}$$

In particular, equality is attained whenever  $\tilde{\nu} \in \mathcal{C}_{\nu'}$ .

The above lemma is the single-site mean-field version of the finite-volume variational principle for lattice spin models applied to the CFLM. This explains the presence of  $\nu'$  in the relative entropy and the quenched free energy, i.e., the third term on the right-hand side of inequality (26). The first two terms on the right of (26) are equal to the negative of the  $\tilde{\nu}$  average of the single-site Hamiltonian for the CFLM since  $\Phi^{(1)}$  is linear in its second argument. We can also deduce similar mean-field variational principle for the initial mean-field model, namely,

$$S(\nu|\alpha) \geq -\Phi^{(1)}(\nu, \nu) - \log \int \alpha(d\hat{\sigma}_i) e^{-\Phi^{(1)}(\nu, \delta_{\hat{\sigma}_i})}. \tag{27}$$

**Proof of Lemma 3.6:** For any  $\tilde{\nu} \in M_{\nu'}$ , the expression on the right-hand side of (26) becomes

$$\begin{aligned} & -\Phi^{(1)}(\pi_1 \tilde{\nu}, \pi_1 \tilde{\nu}) + \tilde{\nu}[\log k(\cdot, \cdot)] - \int \nu'(d\eta_i) \log \int \alpha(d\sigma_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\sigma_i})} k(\sigma_i, \eta_i) \\ &= \int_{S \times S'} \tilde{\nu}(d\xi_i) \log \left\{ \frac{e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\sigma_i})} k(\sigma_i, \eta_i)}{\int \alpha(d\sigma_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\sigma_i})} k(\sigma_i, \eta_i)} \right\} = \int_{S \times S'} \tilde{\nu}(d\xi_i) \log \left\{ \frac{d\gamma_{\nu'}}{d(\alpha \otimes \nu')}(\xi_i|\tilde{\nu}) \right\}. \end{aligned} \tag{28}$$

The proof then follows by showing that

$$S(\tilde{\nu}|\alpha \otimes \nu') - \int_{S \times S'} \tilde{\nu}(d\xi_i) \log \left\{ \frac{d\gamma_{\nu'}}{d(\alpha \otimes \nu')}(\xi_i|\tilde{\nu}) \right\} \geq 0. \tag{29}$$

The case for measures  $\tilde{\nu} \in M_{\nu'}$  with  $S(\tilde{\nu}|\alpha \otimes \nu') = \infty$  is trivial since we get strict inequality by the boundedness properties of  $\Phi$  and  $k$ .

Now for the case of  $\tilde{\nu} \in M_{\nu'}$  with  $S(\tilde{\nu}|\alpha \otimes \nu') < \infty$ , we obtain

$$\begin{aligned} & \int_{S \times S'} \tilde{\nu}(d\xi_i) \log \left\{ \frac{d\tilde{\nu}}{d(\alpha \otimes \nu')}(\xi_i) \right\} - \int_{S \times S'} \tilde{\nu}(d\xi_i) \log \left\{ \frac{d\gamma_{\nu'}}{d(\alpha \otimes \nu')}(\xi_i|\tilde{\nu}) \right\} \\ &= \int_{S \times S'} \tilde{\nu}(d\xi_i) \log \left\{ \frac{d\tilde{\nu}}{d\gamma_{\nu'}(\cdot|\tilde{\nu})}(\xi_i) \right\} = S(\tilde{\nu}|\gamma_{\nu'}(\cdot|\tilde{\nu})) \geq 0, \end{aligned} \tag{30}$$

since  $S(\tilde{\nu}|\gamma_{\nu'}(\cdot|\tilde{\nu})) \geq 0$  with equality holding only when  $\tilde{\nu} = \gamma_{\nu'}(\cdot|\tilde{\nu})$ . This concludes the proof.  $\square$

We now state as our next result a theorem concerning some functional  $\Psi_{\nu'}$  on  $M_{\nu'}$  which is dominated by  $J_{\nu'}$  and coincides with  $J_{\nu'}$  on  $\mathcal{C}_{\nu'}$ .

**Theorem 3.7:** *For any given  $\nu' \in \mathcal{P}(S')$ , the function  $J_{\nu'}$  satisfies*

$$J_{\nu'}(\tilde{\nu}) \geq \Psi_{\nu'}(\tilde{\nu}), \tag{31}$$

where

$$\Psi_{\nu'}(\tilde{\nu}) = \Phi(\pi_1 \tilde{\nu}) - \Phi^{(1)}(\pi_1 \tilde{\nu}, \pi_1 \tilde{\nu}) - \int \nu'(d\eta_i) \log \int \alpha(d\hat{\sigma}_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\hat{\sigma}_i})} k(\hat{\sigma}_i, \eta_i).$$

In particular,  $J_{\nu'}$  coincides with  $\Psi_{\nu'}$  on  $\mathcal{C}_{\nu'}$  and if  $\Phi$  is homogeneous of degree  $p$ , then  $\Psi_{\nu'}$  becomes

$$\Psi_{\nu'}(\tilde{\nu}) = (1 - p)\Phi(\pi_1 \tilde{\nu}) - \int \nu'(d\eta_i) \log \int \alpha(d\hat{\sigma}_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}, \delta_{\hat{\sigma}_i})} k(\hat{\sigma}_i, \eta_i). \tag{32}$$

**Proof:** The expression for  $\Psi_{\nu'}$  and the inequality in (31) follow by substituting the lower bound on  $S(\cdot|\alpha \otimes \nu')$  in Lemma 3.6 into the expression for  $J_{\nu'}$  in (20) of Proposition 3.2. Furthermore, if  $\Phi$  is homogeneous of degree  $p$ , we then have

$$\Phi^{(1)}(\pi_1 \tilde{\nu}, \pi_1 \tilde{\nu}) = \left. \frac{d}{dt} \Phi(\pi_1 \tilde{\nu} + t\pi_1 \tilde{\nu}) \right|_{t=0} = \left. \frac{d}{dt} (1+t)^p \Phi(\pi_1 \tilde{\nu}) \right|_{t=0} = p\Phi(\pi_1 \tilde{\nu}), \tag{33}$$

and putting this into the expression for  $\Psi_{\nu'}$  in (31) yields the desired expression in (32).  $\square$

**Corollary 3.8:** The transformed LDP rate function  $J'$  now becomes

$$J'(\nu') = S(\nu'|\alpha') + \Phi'(\nu') - \text{const}, \tag{34}$$

where

$$\Phi'(\nu') = \inf_{\tilde{\nu} \in \mathcal{C}_{\nu'}} \Psi_{\nu'}(\tilde{\nu}).$$

**Proof:** The above expression for the transformed rate function  $J'$  is a consequence of the fact that  $J_{\nu'}$  coincides with  $\Psi_{\nu'}$  on  $\mathcal{C}_{\nu'}$  and  $\mathcal{C}_{\nu'}$  contains the minimizers of  $J_{\nu'}$ .  $\square$

As explained in the above,  $\Phi'$  is the interaction for the transformed system arising from the initial system described by  $\Phi$  and subjected to the sitewise transformations governed by  $k$ .

### 2. Examples

Take  $\Phi$  to be an Ising mean-field interaction (i.e.,  $S = \{+1, -1\}$ ) given by

$$\Phi(m) = -\frac{\beta}{p} m^p, \tag{35}$$

where  $m \in [-1, 1]$  and  $p \geq 1$ . Here the reason for using  $m$  instead of probability measures on  $S$  is that the probability measures on  $S$  are uniquely determined by  $m$ , i.e., each  $m \in [-1, 1]$  can uniquely be associated with a probability measure (say,  $\nu$ ) on  $S$  given by  $\nu(\sigma_i) = [(1+m)/2]\delta_{+1}(\sigma_i) + [(1-m)/2]\delta_{-1}(\sigma_i)$ . The expectation with respect to  $\nu$  then gives rise to  $m$ .

We take  $k(\sigma_i, \eta_i) = p_t(\sigma_i, \eta_i)$  to be the transition probabilities [i.e.,  $p_t(\sigma_i, \eta_i)$  is the probability of starting with  $\sigma_i$  at site  $i$  and observing  $\eta_i$  after  $t$  time units] for rate one sitewise independent spin-flip dynamics on  $S$ .<sup>25</sup> Here both  $S$  and  $S'$  are the same, and the *a priori* measures  $\alpha = \alpha' = \frac{1}{2}(\delta_{+1} + \delta_{-1})$ . More precisely,  $p_t(\sigma_i, \eta_i)$  is given by

$$p_t(\sigma_i, \eta_i) = \frac{e^{\sigma_i \eta_i h_t}}{2 \cosh(h_t)}, \tag{36}$$

where

$$h_t = \frac{1}{2} \log \frac{1 + e^{-2t}}{1 - e^{-2t}}.$$

As pointed out above, we will denote by  $m' \in [-1, 1]$  the expected values of the probability measures on the transformed single-site space. We will also write  $m$  for the expected values of the first marginals of probability measures on  $S \times S$  with fixed second marginal having  $m'$  as its mean.

Then  $\Psi_{m'}$  for this setup becomes

$$\begin{aligned} \Psi_{m'}(m) &= \frac{(p-1)\beta}{p} m^p - \frac{1+m'}{2} \log \{ \cosh(\beta m^{(p-1)} + h_t) \} - \frac{1-m'}{2} \log \{ \cosh(\beta m^{(p-1)} - h_t) \} \\ &\quad + \log(2 \cosh(h_t)). \end{aligned} \tag{37}$$

Consequently, this form of  $\Psi_{m'}$  gives rise to the mean-field equation

$$m = \frac{1+m'}{2} \tanh(\beta m^{(p-1)} + h_t) + \frac{1-m'}{2} \tanh(\beta m^{(p-1)} - h_t) \tag{38}$$

after differentiating  $\Psi_{m'}$  and setting the derivative to zero.

**Remark:** In the case  $p=2$ ,  $\Psi$  is the Hubbard–Stratonovitch potential function.<sup>25</sup> The unicity of global minimizers of this potential function played a crucial role in determining the Gibbs and non-Gibbs properties of the corresponding transformed system studied in Ref. 25.

In Ref. 25 the derivation of  $\Psi$  was based on a technique that exploits the quadratic nature of the interaction  $\Phi$ . This technique cannot be used to derive  $\Psi$  for nonquadratic interactions. Thus our approach of deriving  $\Psi$  via the machinery of large deviations theory is more robust since it is applicable to both quadratic and nonquadratic interactions in more general (possibly continuous) spin spaces.

We now return to the discussion of Gibbsianness for transforms of initial Gibbsian mean-field models as introduced above.

### B. Gibbsianness for transforms of mean-field models

In this subsection we study the Gibbs properties of the transformed measures  $\mu'$  (18) introduced in Sec. III A. Before we formalize our discussion let us fix some notations that we shall use in our study. For each  $N \geq 2$  and  $1 \leq n < N$  we denote by  $\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}]$  the joint system in  $V_{N-n}$  when the second-layer spins are constrained to a given configuration  $\bar{\eta} \in \Omega'$ , i.e.,  $\bar{\eta}_{V_{N-n}}$  is the projection of  $\bar{\eta}$  onto  $S'^{N-n}$ . As we pointed out before, a representative  $\bar{\eta}_{V_{N-n}}$  of a class of configurations in  $S'^{N-n}$  with the same empirical measure  $L_{N-n}(\bar{\eta}_{V_{N-n}})$  will give rise to the same measure  $\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}]$ . Therefore by fixing  $\bar{\eta}_{V_{N-n}}$  implies that we are restricting attention to only the configurations in a subset of  $S'^{N-n}$  with fixed  $L_{N-n}(\bar{\eta}_{V_{N-n}})$ . Suppose that  $\bar{\eta}_{V_{N-n}}$  is one such representative; we call  $\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}]$  the *restricted constrained first-layer model* (RCFLM) for the corresponding two-layer mean-field model. It is restricted because we are not taking into account the spins in  $V_n$  and constrained since we have frozen the configurations in the second layer to  $\bar{\eta}_{V_{N-n}}$ . More precisely,

$$\begin{aligned} \tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}](d\sigma_{V_{N-n}}) &= \frac{\exp(-N\Phi(\pi_1 \bar{v}_{N,n})) \prod_{i=n+1}^N k(\sigma_i, \bar{\eta}_i) \alpha(d\sigma_i)}{\int \exp(-N\Phi(\pi_1 \hat{v}_{N,n})) \prod_{i=n+1}^N k(\hat{\sigma}_i, \bar{\eta}_i) \alpha(d\hat{\sigma}_i)} \\ &= \frac{\exp(-N\{\Phi(\pi_1 \bar{v}_{N,n}) - \bar{v}_{N,n}[\log k]\}) \prod_{i=n+1}^N \alpha(d\sigma_i)}{\int \exp(-N\{\Phi(\pi_1 \hat{v}_{N,n}) - \hat{v}_{N,n}[\log k]\}) \prod_{i=n+1}^N \alpha(d\hat{\sigma}_i)}, \end{aligned} \tag{39}$$

where  $\bar{v}_{N,n} = [(N-n)/N] L_{N-n}(\bar{\xi}_{V_{N-n}})$ ,  $L_{N-n}(\bar{\xi}_{V_{N-n}}) = [(1/N-n) \sum_{i=n+1}^N \delta_{(\sigma_i, \bar{\eta}_i)}]$ , and  $\hat{\xi}_i = (\hat{\sigma}_i, \bar{\eta}_i)$ .

**Remark:** Suppose  $\nu' \in \mathcal{P}(S')$  is the empirical measure for the configuration  $\bar{\eta}$ . Then for a fixed  $n$ , the sequence of measures  $\bar{\nu}_{N,n}$  under the pushed forwards of  $\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}]$  satisfies a LDP with rate  $N$  and rate function  $\bar{J}_{\nu'} : M_{\nu'} \rightarrow \mathbb{R} \cup \{+\infty\}$  given by

$$\bar{J}_{\nu'}(\tilde{\nu}) = J_{\nu'}(\tilde{\nu}) - \text{const}_{\nu'}, \tag{40}$$

where

$$\text{const}_{\nu'} = \inf_{\tilde{\nu} \in M_{\nu'}} J_{\nu'}(\tilde{\nu}).$$

The validity of the expression for  $\bar{J}_{\nu'}$  lies in the fact that for  $\bar{\eta} \in \Omega'$  constrained to have empirical measure  $\nu'$ , the sequence of empirical measures  $L_{N-n}(\bar{\xi}_{V_{N-n}})$  converges weakly in the  $N$ -limit to an element in  $M_{\nu'}$ . For any of such measures  $\tilde{\nu}$  for which  $\tilde{\nu}(\cdot | \eta_i) \ll \alpha$  (for  $\nu'$  almost all  $\eta_i \in S'$ ),

$$\frac{d\tilde{\nu}}{d\alpha}(\sigma_i, d\eta_i) = \nu'(d\eta_i) f_{\nu'}(\sigma_i | \eta_i). \tag{41}$$

Therefore, the relative entropy of such probability measures  $\tilde{\nu}$  with respect to  $\alpha$  then becomes

$$S(\tilde{\nu} | \alpha) = \int \nu'(d\eta_i) \int f_{\nu'}(\sigma_i | \eta_i) \log f_{\nu'}(\sigma_i | \eta_i) \alpha(d\sigma_i). \tag{42}$$

Our next result in this subsection concerns a representation of the finite-volume transformed conditional distributions  $\mu'_N(\cdot | \bar{\eta}_{V_{N-n}})$  in terms of the RCFLM  $\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}]$ .

**Lemma 3.9:** *Let  $N, n$ , and  $\bar{\eta}_{V_{N-n}}$  be as above. Then the finite-volume conditional distribution  $\mu'_N(\cdot | \bar{\eta}_{V_{N-n}})$  for the transformed system has the form*

$$\mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}}) = \frac{\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}] \left[ \prod_{i=1}^n \int_S e^{-\Phi^{(1)}(\pi_1 \bar{\nu}_{N,n}, \delta_{\sigma_i} + o(1/N))} k(\sigma_i, \eta_i) \alpha(d\sigma_i) \alpha'(d\eta_i) \right]}{\tilde{\mu}_{N-n}[\bar{\eta}_{V_{N-n}}] \left[ \prod_{i=1}^n \int_S e^{-\Phi^{(1)}(\pi_1 \bar{\nu}_{N,n}, \delta_{\sigma_i} + o(1/N))} \alpha(d\sigma_i) \right]}. \tag{43}$$

**Proof:** Note from the definition of the transformed system that we can write  $\mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}})$  as

$$\mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}}) = \frac{\prod_{j=n+1}^N \int_S \alpha(d\sigma_j) k(\bar{\xi}_j) \prod_{i=1}^n \int_S \alpha(d\sigma_i) \alpha'(d\eta_i) e^{-N\Phi(\pi_1 L_N(\bar{\xi}_{V_N}))} k(\hat{\xi}_i)}{\prod_{j=n+1}^N \int_S \alpha(d\sigma_j) k(\bar{\xi}_j) \prod_{i=1}^n \int_{S \times S'} \alpha(d\sigma_i) \alpha'(d\hat{\eta}_i) e^{-N\Phi(\pi_1 L_N(\bar{\xi}_{V_N}))} k(\hat{\xi}_i)}, \tag{44}$$

where the joint configuration  $\bar{\xi}_{V_N}$  is such that  $\bar{\xi}_{V_{N-n}} = (\sigma_i, \bar{\eta}_i)_{i \in V_{N-n}}$  and  $\bar{\xi}_{V_n} = (\sigma_i, \eta_i)_{i \in V_n}$ , and  $\hat{\xi}_i = (\sigma_i, \hat{\eta}_i)$ . Now by writing the joint empirical measure as

$$L_N(\bar{\xi}_{V_N}) = \frac{N-n}{N} L_{N-n}(\bar{\xi}_{V_{N-n}}) + \frac{n}{N} L_n(\hat{\xi}_{V_n}),$$

and adding and subtracting  $N\Phi([(N-n)/N] \pi_1 L_{N-n}(\bar{\xi}_{V_{N-n}}))$  from the exponent  $N\Phi(\pi_1 L_N(\bar{\xi}_{V_N}))$ , we obtain

$$N\Phi(\pi_1 L_N(\bar{\xi}_{V_N})) = N\Phi\left(\frac{N-n}{N} \pi_1 L_{N-n}(\bar{\xi}_{V_{N-n}})\right) + \sum_{i=1}^n \Phi^{(1)}\left(\frac{N-n}{N} \pi_1 L_{N-n}(\bar{\xi}_{V_{N-n}}), \delta_{\sigma_i}\right) + o\left(\frac{1}{N}\right), \tag{45}$$

where  $o(1/N)$  is a result of (1) of Definition 2.1 where we have taken  $t=1/N$ . Finally, by putting this expression of  $N\Phi(\pi_1 L_N(\bar{\xi}_{V_N}))$  (45) into the expression of  $\mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}})$  (44) and multiplying the resulting expression by

$$\frac{\prod_{j=n+1}^N \int \alpha(d\sigma_j) \exp(-N\Phi(\pi_1 \bar{v}_{N,n})) k(\sigma_j, \bar{\eta}_j)}{\prod_{j=n+1}^N \int \alpha(d\sigma_j) \exp(-N\Phi(\pi_1 \bar{v}_{N,n})) k(\sigma_j, \bar{\eta}_j)}, \tag{46}$$

we conclude the proof of the lemma. □

We now state the infinite-volume ( $N \rightarrow \infty$ ) version of Lemma 3.9. A sufficient condition for the existence of the finite-volume conditional distributions with infinite-volume  $\eta$ -conditioning is provided. This sufficient condition is the unicity of the global minimizers of the function  $J_{\nu'}$ ,  $\nu' \in \mathcal{P}(S')$ .

**Theorem 3.10:** *For each  $\nu' \in \mathcal{P}(S')$ , let  $J_{\nu'}: M_{\nu'} \rightarrow \mathbb{R} \cup \{+\infty\}$  be as defined in (20). Suppose further that for a given  $\nu' \in \mathcal{P}(S')$   $J_{\nu'}$  has a unique global minimizer  $\tilde{\nu}^* \in M_{\nu'}$ .*

(I) Then

$$\gamma'_n(d\eta_{V_n} | \nu') = \lim_{N \rightarrow \infty} \mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}}) = \prod_{i \in V_n} \gamma'_1(d\eta_i | \nu'), \tag{47}$$

with

$$\gamma'_1(d\eta_i | \nu') = \frac{\int_S \alpha(d\sigma_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i})} k(\sigma_i, \eta_i) \alpha'(d\eta_i)}{\int_S \alpha(d\sigma_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i})}}.$$

(II) If  $J_{\lambda'}$  has a unique global minimizer  $\tilde{\lambda}^*$  for all  $\lambda'$  in a weak neighborhood of  $\nu'$ , then  $\gamma'_n(d\eta_{V_n} | \nu')$  is weakly continuous at  $\nu'$  as a function of the conditioning  $\nu' \in \mathcal{P}(S')$ .

**Remark:**

- (1) The  $\nu'$  dependence of the expression for  $\gamma'_1(d\sigma'_1 | \nu')$  is hidden in  $\Phi^{(1)}$  via the probability measure  $\pi_1 \tilde{\nu}^*$ .
- (2) Theorem 3.10 provides a sufficient condition for the transformed system to be Gibbs (in the sense of Definition 2.2), namely, the unicity of global minimizers of  $J$ . Thus the problem of determining whether the transformed system is Gibbs or not is then translated into the corresponding problem of studying the global minimizers of  $J$ .

**Proof of Theorem 3.10:**

- (I) The proof follows from the representation of the finite-volume conditional distribution  $\mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}})$  given in Lemma 3.9 and the hypothesis that the function  $J_{\nu'}$  has a unique global minimizer because the leading term in the large  $N$  asymptotic of  $\mu'_N(d\eta_{V_n} | \bar{\eta}_{V_{N-n}})$  is governed by the global minimizers of  $J_{\nu'}$ , i.e., the global minimizers of the rate function for the RCFLM  $\mu_{N-n}[\eta_{V_{N-n}}]$ .
- (II) The continuity property of the transformed kernel  $\gamma'_1$  trivially follows from the continuity property of  $\Phi^{(1)}$  in its first argument, hence the proof that the transformed system is Gibbs

whenever the functional  $J_{\nu'}$  has a unique global minimizer uniformly in  $\nu'$ .

□

#### IV. GIBBSIANNES OF TRANSFORMED SYSTEMS AND THE CONTRACTION MAP THEOREM

This section is devoted for studying the minimizers of the function  $J$  for some special class of initial interactions  $\Phi$ . Up to this point all topological considerations have been with respect to the weak topology, i.e., the weak topology is sufficient to study Gibbs measures and Gibbs properties of transforms of Gibbs measures for mean-field models. We now consider another topology on the spaces of measures which is stronger than the weak topology. This topology is the one induced by the total variational metric. Continuity in this new topology implies the continuity with respect to the weak topology. All topological considerations for the interactions we consider in this section shall be with respect to the variational topology. Additionally, we also impose further smoothness requirements on the initial interactions  $\Phi$  other than those given in Definition 2.1. All these restrictions on the interactions are required to derive explicit continuity estimates on the CFLPKs and consequently on the transformed kernels.

To be precise, we consider interactions  $\Phi$  that are given by

$$\Phi(\nu) = F(\nu[g_1], \dots, \nu[g_l]), \quad (48)$$

where  $g_i$  are some fixed bounded nonconstant real-valued measurable functions defined on  $S$ ,  $l \geq 1$  and  $F: \mathbb{R}^l \rightarrow \mathbb{R}$  is some twice continuously differentiable function (e.g., if  $F$  is a polynomial). In the following we will write  $g = (g_1, \dots, g_l)$  and  $\nu[g] = (\nu[g_1], \dots, \nu[g_l])$ . By setting  $m_j = m_j(\nu) = \nu[g_j]$ , we have for this choice of interaction that

$$\Phi^{(1)}(\nu, \delta_{\sigma_i}) = \sum_{j=1}^l F_j(\nu[g_1], \dots, \nu[g_l]) g_j(\sigma_i), \quad (49)$$

where  $F_j(m) = (\partial/\partial m_j)F(m)$  and  $m = m(\nu) = \nu[g]$ . We also set  $F_{ju}(m) = (\partial^2/\partial m_j \partial m_u)F(m)$ . Additionally, we assume that  $g$  is a Lipschitz function from  $S$  to  $\mathbb{R}^l$ , with Lipschitz norm

$$\|g\|_{d,2} = \sup_{\sigma_i \neq \bar{\sigma}_i} \frac{\|g(\sigma_i) - g(\bar{\sigma}_i)\|_2}{d(\sigma_i, \bar{\sigma}_i)}, \quad (50)$$

where  $d$  is the metric on  $S$ . We also denote by  $\delta(g)$  the sum of the oscillations of the components of  $g$ , i.e.,

$$\delta(g) = \sum_{j=1}^l \delta(g_j). \quad (51)$$

For any  $g$  satisfying the above conditions we set

$$D_g = \overline{\{\nu[g]: \nu \in \mathcal{P}(S)\}}. \quad (52)$$

Note that  $D_g$  is compact subset of  $\mathbb{R}^l$  by the boundedness of  $g$ . In the sequel we will write  $\|\partial^2 F\|_{\max, \infty}$  for the supremum of the matrix max norm of the Hessian  $\partial^2 F$ , i.e.,

$$\|\partial^2 F\|_{\max, \infty} = \sup_{m \in D_g} \|\partial^2 F(m)\|_{\max}, \quad (53)$$

where

$$\|\partial^2 F(m)\|_{\max} = \max_{1 \leq i, j \leq l} |F_{ij}(m)|.$$

Furthermore, we also set

$$\delta_{F,g} = \sup_{m \in D_g} \sup_{\sigma_i, \bar{\sigma}_i \in S} \left| \sum_{j=1}^l F_j(m)(g_j(\sigma_i) - g_j(\bar{\sigma}_i)) \right|. \tag{54}$$

Up to this point one may wonder whether the class of interactions we are considering in this section has any physical relevance. Indeed, it contains important mean-field interactions such as the Curie–Weiss interactions, liquid crystal interactions, and sums of “*p*-spin” interactions that have featured prominently in the literature.

**A. Lipschitz continuity of the CFLPK and Gibbsianness of transformed system**

We have already seen from the remark below Eq. (25) that the CFLPK  $\gamma$  is weakly continuous. In this subsection we show, however, that the CFLPK is Lipschitz continuous with respect to the variational metric (defined below). We write

$$\|\nu - \bar{\nu}\| = \sup_{|\varphi| \leq 1} |\nu(\varphi) - \bar{\nu}(\varphi)| = \sup_{\bar{\varphi}} \frac{|\nu(\bar{\varphi}) - \bar{\nu}(\bar{\varphi})|}{\delta(\bar{\varphi})}, \tag{55}$$

where

$$\delta(\bar{\varphi}) = \sup_{\sigma_i \neq \bar{\sigma}_i} |\bar{\varphi}(\sigma_i) - \bar{\varphi}(\bar{\sigma}_i)|,$$

for the variational distance between the probability measures  $\nu$  and  $\bar{\nu}$  where the supremums are, respectively, taken over all measurable real-valued functions  $\varphi$  with  $|\varphi| \leq 1$  and bounded nonconstant measurable real-valued functions on  $S$ . The variational distance can also be defined by the following consideration: The signed measure  $\nu - \bar{\nu}$  has, respectively,  $(\nu - \bar{\nu})^+$  and  $(\nu - \bar{\nu})^-$  as the positive and negative parts of its Jordan decomposition. However, the fact that  $(\nu - \bar{\nu})(S) = 0$  implies that  $(\nu - \bar{\nu})^+(S) = (\nu - \bar{\nu})^-(S)$ , leading to the definition of the variational distance between  $\nu$  and  $\bar{\nu}$  as one-half of the total variation of  $(\nu - \bar{\nu})$ , i.e.,

$$\|\nu - \bar{\nu}\| = (\nu - \bar{\nu})^+(S) = (\nu - \bar{\nu})^-(S). \tag{56}$$

Before we state our first result in this section let us fix further notations. We set

$$C(F, g) = 2 \|\partial^2 F\|_{\max, \infty} \delta(g) \|g\|_{d,2} \exp\left(\frac{\delta_{F,g}}{2}\right) \quad \text{and}$$

$$\rho_\alpha(k) = \sup_{\eta_i \in S'} \inf_{a_i \in S} \left( \int_S d^2(\sigma_i, a_i) k(\sigma_i, \eta_i) \alpha(d\sigma_i) \right)^{1/2}. \tag{57}$$

**Theorem 4.1:** *For any  $\nu' \in \mathcal{P}(S)$  and each pair  $\tilde{\nu}_1, \tilde{\nu}_2 \in M_{\nu'}$ , the CFLPK satisfies*

$$\|\gamma_{\nu'}(\cdot | \tilde{\nu}_1) - \gamma_{\nu'}(\cdot | \tilde{\nu}_2)\| \leq L \|\tilde{\nu}_1 - \tilde{\nu}_2\|, \tag{58}$$

where

$$L = L(F, g, k) = C(F, g) \rho_\alpha(k).$$

The above theorem says that for each  $\nu' \in \mathcal{P}(S')$ , the CFLPK  $\gamma_{\nu'}$  is Lipschitz continuous on  $M_{\nu'}$  with Lipschitz constant  $L$ .

**Remark:**

- (1) The quantity  $\rho_\alpha(k)$  is the (metric-space version of) standard deviation of the single-site “posterior distribution”  $K(d\sigma_i | \eta_i)$  when we take supremum over the possible observations  $\eta$ . So, it describes the worst- $\eta$  scenario of the typical size of fluctuations in the initial configurations which have led to  $\eta$ . The constant  $L$  factorizes into two constants reflecting the idea of “nature  $C(F, g)$  versus nurture  $\rho_\alpha(k)$ .”

(2) Set

$$\rho_\alpha = \inf_{a_i \in S} \left( \int \alpha(d\sigma_i) d^2(\sigma_i, a_i) \right)^{1/2}, \tag{59}$$

i.e.,  $\rho_\alpha$  is the metric-space version of the standard deviation of  $\alpha$ . Then the initial kernel  $\gamma_1$  for the interactions considered in this section is also Lipschitz continuous, i.e., for any pair  $\nu_1, \nu_2 \in \mathcal{P}(S)$  we have

$$\|\gamma_1(\cdot|\nu_1) - \gamma_1(\cdot|\nu_2)\| \leq \hat{L} \|\nu_1 - \nu_2\|, \tag{60}$$

where

$$\hat{L} = C(F, g)\rho_\alpha.$$

$\hat{L}$  is the ‘‘Dobrushin’s constant’’ for the initial mean-field model.

**Proof of Theorem 4.1:** Take a measurable map  $f: S \times S' \rightarrow \mathbb{R}$ , with  $|f| \leq 1$ . Also for any pair  $\tilde{\nu}_1, \tilde{\nu}_2 \in M_{\nu'}$  and any  $0 \leq s \leq 1$  we define  $\tilde{\nu}_s := s\tilde{\nu}_1 + (1-s)\tilde{\nu}_2$ . Then we have

$$|\gamma_{\nu'}(f|\tilde{\nu}_1) - \gamma_{\nu'}(f|\tilde{\nu}_2)| = \left| \int \nu'(d\eta_i) \int \alpha(d\sigma_i) f(\sigma_i, \eta_i) \int_0^1 ds \frac{d}{ds} h_{\tilde{\nu}_s}(\sigma_i, \eta_i) \right|, \tag{61}$$

where

$$h_{\tilde{\nu}_s}(\sigma_i, \eta_i) = \frac{e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}_s, \delta_{\sigma_i})} k(\sigma_i, \eta_i)}{\int e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}_s, \delta_{\sigma_i})} k(\sigma_i, \eta_i) \alpha(d\sigma_i)}.$$

We also set  $\lambda_s[\eta_i](d\sigma_i) = h_{\tilde{\nu}_s}(\sigma_i, \eta_i) \alpha(d\sigma_i)$ . Now using the form of the interactions considered in this section, it is not hard to deduce that

$$\frac{d}{ds} \Phi^{(1)}(\pi_1 \tilde{\nu}_s, \delta_{\sigma_i}) = \sum_{j=1}^l \sum_{u=1}^l F_{ju}(m(\pi_1 \tilde{\nu}_s)) g_j(\sigma_i) \int \pi_1(\tilde{\nu}_1 - \tilde{\nu}_2)(d\bar{\sigma}_i) g_u(\bar{\sigma}_i). \tag{62}$$

This and further computations yield the following expression for  $(d/ds)h_{\tilde{\nu}_s}$ :

$$\frac{d}{ds} h_{\tilde{\nu}_s}(\sigma_i, \eta_i) = - \sum_{j=1}^l \sum_{u=1}^l F_{ju}(m(\pi_1 \tilde{\nu}_s)) \psi_u h_{\tilde{\nu}_s}(\sigma_i, \eta_i) (g_j(\sigma_i) - \lambda_s[\eta_i](g_j)), \tag{63}$$

where

$$\psi_u = \int \pi_1(\tilde{\nu}_1 - \tilde{\nu}_2)(d\bar{\sigma}_i) g_u(\bar{\sigma}_i).$$

Observe from (55) that

$$|\psi_u| = \delta(g_u) \frac{\left| \int \nu'(d\eta_i) \int (\tilde{\nu}_1(d\sigma_i|\eta_i) - \tilde{\nu}_2(d\sigma_i|\eta_i)) g_u(\sigma_i) \right|}{\delta(g_u)} \leq \delta(g_u) \|\tilde{\nu}_1 - \tilde{\nu}_2\|. \tag{64}$$

Putting all these together we arrive at

$$|\gamma_{\nu'}(f|\tilde{\nu}_1) - \gamma_{\nu'}(f|\tilde{\nu}_2)| \leq \|\tilde{\nu}_1 - \tilde{\nu}_2\| \sum_{j=1}^l \sum_{u=1}^l \delta(g_u) \int_0^1 ds |F_{ju}(m(\pi_1 \tilde{\nu}_s))| \int \nu'(d\eta_i) \varphi(s, \eta_i), \quad (65)$$

where

$$\varphi(s, \eta_i) = \int \alpha(d\sigma_i) h_{\tilde{\nu}_s}(\sigma_i, \eta_i) |g_j(\sigma_i) - \lambda_s[\eta_i](g_j)|.$$

By adding and subtracting  $g_j(a_i)$  (for any arbitrary  $a_i \in S$ ) from the term  $g_j(\sigma_i) - \lambda_s[\eta_i](g_j)$  in the definition of  $\varphi(s, \eta_i)$  and applying the triangle inequality we arrive at the following:

$$\varphi(s, \eta_i) \leq 2 \int \alpha(d\sigma_i) h_{\tilde{\nu}_s}(\sigma_i, \eta_i) |g_j(\sigma_i) - g_j(a_i)|. \quad (66)$$

Further, it follows from Hölder's inequality that

$$\int \alpha(d\sigma_i) h_{\tilde{\nu}_s}(\sigma_i, \eta_i) |g_j(\sigma_i) - g_j(a_i)| \leq \left( \int \alpha(d\sigma_i) h_{\tilde{\nu}_s}(\sigma_i, \eta_i) (g_j(\sigma_i) - g_j(a_i))^2 \right)^{1/2}. \quad (67)$$

Now by replacing  $|F_{ju}(m(\pi_1 \tilde{\nu}_s))|$  with  $\|\partial^2 F\|_{\max, \infty}$  in (65) and using the fact that the square root function is concave, we obtain

$$|\gamma_{\nu'}(f|\tilde{\nu}_1) - \gamma_{\nu'}(f|\tilde{\nu}_2)| \leq 2\|\tilde{\nu}_1 - \tilde{\nu}_2\| \delta(g) \|\partial^2 F\|_{\max, \infty} \int_0^1 ds \int \nu'(d\eta_i) \chi(s, \eta_i, a_i), \quad (68)$$

where

$$\chi(s, \eta_i, a_i) = \left( \int \alpha(d\sigma_i) h_{\tilde{\nu}_s}(\sigma_i, \eta_i) \sum_{j=1}^l (g_j(\sigma_i) - g_j(a_i))^2 \right)^{1/2}.$$

Note further that the  $a_i$  appearing in  $\chi$  is chosen independent of all the parameters in the model, so taking the infimum over  $a_i$  will have no influence on our estimates. In view of this observation, replacing  $\chi(s, \eta_i, a_i)$  with  $\inf_{a_i \in S} \chi(s, \eta_i, a_i)$  will have no effect on the inequality in (68). Furthermore, it follows from the Lipschitz property of  $g$  and the fact that  $h_{\tilde{\nu}_s}(\sigma_i, \eta_i) \leq e^{\delta F_s k} k(\sigma_i, \eta_i)$  that

$$\inf_{a_i \in S} \chi(s, \eta_i, a_i) \leq \|g\|_{d, 2} e^{\delta F_s g/2} \sup_{\eta_i \in S'} \inf_{a_i \in S} \left( \int \alpha(d\sigma_i) k(\sigma_i, \eta_i) d^2(\sigma_i, a_i) \right)^{1/2}. \quad (69)$$

Putting the bound in (69) into the bound in (68) yields the desired result. □

Observe from Theorem 4.1 that if the constant  $L < 1$ , then  $\gamma_{\nu'}$  defines a contraction map from  $M_{\nu'}$  to itself. This is because the variational distance turns the set  $M_{\nu'} = \pi_2^{-1}(\{\nu'\})$  into a complete metric space by continuity of the map  $\pi_2$  under the variational topology. Thus the CFLPK admits a unique consistent probability measure and consequently the existence of a unique global minimizer for  $J$ . since the minimizers of  $J$ . are contained in the set of consistent probability measures for the CFLPK. This then implies that the transformed system is Gibbs.

The next item on our list of tasks is the investigation of how (in the regime  $L < 1$ ) the unique consistent probability measure  $\tilde{\nu}^*$  for  $\gamma_{\nu'}$  behaves with respect to  $\nu' \in \mathcal{P}(S')$ . Indeed we show in the proposition below that  $\tilde{\nu}^*$  depends continuously on  $\nu'$ .

**Proposition 4.2:** *Suppose the constant  $L < 1$  ; then under the variational metric the unique consistent probability measure  $\tilde{\nu}^*$  for  $\gamma_{\nu'}$  is Lipschitz continuous with respect to  $\nu'$  and has the Lipschitz norm  $L_1 = 4L$ .*

**Remark:** The constant  $L$  here is comparable to the uniform bound on the Dobrushin constant  $c'[\eta]$  for the RCFLM considered in Ref. 24. Due to this, it is of interest to obtain the Lipschitz constant with the factor  $1/(1-L)$ , reminiscent of the upper bound on the row sums of the Do-

brushin matrix  $D$ . Thus for  $L < 1$  we can also have the constant  $\bar{L}_1$  given by

$$\bar{L}_1 = \frac{L_1}{1 - L}, \tag{70}$$

with

$$L_1 < \bar{L}_1.$$

**Proof of Proposition 4.2:** Let  $\nu'_1, \nu'_2 \in \mathcal{P}(S')$ ; then the assertion of the proposition follows by showing that

$$\|\gamma_{\nu'_1} - \gamma_{\nu'_2}\| \leq L_1 \|\nu'_1 - \nu'_2\| \tag{71}$$

since if  $\tilde{\nu}_i^*$  is the unique consistent probability measure for  $\gamma_{\nu'_i}$ , then  $\tilde{\nu}_i^* = \gamma_{\nu'_i}(\cdot | \tilde{\nu}_i^*)$  for  $i=1, 2$ . Observe for any measurable function on  $S \times S'$  with  $|f| \leq 1$  that

$$\begin{aligned} |\gamma_{\nu'_1}(f | \tilde{\nu}_1^*) - \gamma_{\nu'_2}(f | \tilde{\nu}_2^*)| &= \left| \int (\nu'_1 - \nu'_2)(d\eta_i) \alpha(d\sigma_i) f(\sigma_i, \eta_i) (h_{\tilde{\nu}_1^*}(\sigma_i, \eta_i) - h_{\tilde{\nu}_2^*}(\sigma_i, \eta_i)) \right| \\ &\leq \left| \int (\nu'_1 - \nu'_2)^+(d\eta_i) \alpha(d\sigma_i) f(\sigma_i, \eta_i) (h_{\tilde{\nu}_1^*}(\sigma_i, \eta_i) - h_{\tilde{\nu}_2^*}(\sigma_i, \eta_i)) \right| \\ &\quad + \left| \int (\nu'_1 - \nu'_2)^-(d\eta_i) \alpha(d\sigma_i) f(\sigma_i, \eta_i) (h_{\tilde{\nu}_1^*}(\sigma_i, \eta_i) - h_{\tilde{\nu}_2^*}(\sigma_i, \eta_i)) \right|, \end{aligned} \tag{72}$$

where  $(\nu'_1 - \nu'_2)^+$  and  $(\nu'_1 - \nu'_2)^-$  are, respectively, the positive and negative parts of the Jordan decomposition of the signed measure  $\nu'_1 - \nu'_2$  and  $h_{\tilde{\nu}}$  is as given in (61). It follows from the definition of the variational distance between two probability measures that

$$|\gamma_{\nu'_1}(f | \tilde{\nu}_1^*) - \gamma_{\nu'_2}(f | \tilde{\nu}_2^*)| \leq 2 \|\nu'_1 - \nu'_2\| \sup_{\eta_i \in S'} \int \alpha(d\sigma_i) |h_{\tilde{\nu}_1^*}(\sigma_i, \eta_i) - h_{\tilde{\nu}_2^*}(\sigma_i, \eta_i)| \tag{73}$$

since we have chosen  $f$  to be such that  $|f| \leq 1$ . Now we proceed by setting  $\tilde{\nu}_s^* = s\tilde{\nu}_1^* + (1-s)\tilde{\nu}_2^*$ , with  $0 \leq s \leq 1$ , and observing that

$$\begin{aligned} \int \alpha(d\sigma_i) |h_{\tilde{\nu}_1^*}(\sigma_i, \eta_i) - h_{\tilde{\nu}_2^*}(\sigma_i, \eta_i)| &= \int \alpha(d\sigma_i) \left| \int_0^1 ds \frac{d}{ds} h_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) \right| \\ &\leq \int_0^1 ds \int \alpha(d\sigma_i) \left| \frac{d}{ds} h_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) \right|. \end{aligned} \tag{74}$$

However, we know from the proof of Theorem 4.1 that

$$\int_0^1 ds \int \alpha(d\sigma_i) \left| \frac{d}{ds} h_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) \right| \leq C(F, g) \|\tilde{\nu}_1^* - \tilde{\nu}_2^*\| \inf_{a_i \in S} \left( \int \alpha(d\sigma_i) k(\sigma_i, \eta_i) d^2(\sigma_i, a_i) \right)^{1/2}. \tag{75}$$

Now by replacing  $\|\tilde{\nu}_1^* - \tilde{\nu}_2^*\|$  by 2 and substituting the bound in (75) into the inequality in (73), we conclude the proof of the proposition.  $\square$

Having disposed of the continuity estimates for the CFLPK, we now turn our attention to study the corresponding continuity estimates for the single-site kernel  $\gamma'_1$  for the transformed

system in the regime where  $L < 1$  since in this regime the CFLPK has a unique consistent probability measure and consequently provides a sufficient condition for the function  $J$  to have a unique global minimizer as is required by Theorem 3.10.

**B. Continuity estimates for  $\gamma'_1$**

A sufficient condition given in Theorem 3.10 for the existence and continuity of the finite-volume kernels  $\gamma'_n$  for the transformed system is the unicity of the global minimizer of the function  $J_{\nu'}$ . This sufficient condition holds if the CFLPK has a unique consistent probability. Our main result in this subsection is the following theorem concerning the continuity estimate for  $\gamma'_1$ .

**Theorem 4.3:** *Suppose that  $L < 1$  ; then under the variational metric on  $\mathcal{P}(S')$  the single-site kernel  $\gamma'_1(\cdot | \nu')$  is Lipschitz continuous with respect to  $\nu'$  with Lipschitz constant  $L_2$  given by*

$$L_2 = L_1 \hat{L}. \tag{76}$$

**Remark:**

- (1) The Lipschitz constant  $L_2$  for the transformed system factorizes into the product of the Lipschitz constants  $L$  and  $\hat{L}$ , respectively, for the CFLPK  $\gamma$  and the initial kernel  $\gamma$ .
- (2) Observe from the remark below Proposition 4.2 that we can also have the Lipschitz constant  $\bar{L}_2$  given by

$$\bar{L}_2 = \bar{L}_1 \hat{L} = \frac{4L\hat{L}}{1-L}, \tag{77}$$

with

$$L_2 < \bar{L}_2.$$

**Proof of Theorem 4.3:** As usual let us take  $\tilde{\nu}_1^*$  and  $\tilde{\nu}_2^*$  as the unique consistent probability measures for the CFLPK corresponding to  $\nu'_1$  and  $\nu'_2$ , respectively. Again set  $\tilde{\nu}_s^* = s\tilde{\nu}_1^* + (1-s)\tilde{\nu}_2^*$  for  $0 \leq s \leq 1$ . It follows from (47) after taking a measurable map  $f: S' \rightarrow \mathbb{R}$  with  $|f| \leq 1$  and setting

$$\hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) = \frac{k(\sigma_i, \eta_i) e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}_s^*, \delta_{\sigma_i})}}{\int e^{-\Phi^{(1)}(\pi_1 \tilde{\nu}_s^*, \delta_{\hat{\sigma}_i})} \alpha(d\hat{\sigma}_i)} \tag{78}$$

that

$$\gamma'_1(f | \nu'_1) - \gamma'_1(f | \nu'_2) = \int \alpha'(d\eta_i) f(\eta_i) \int \alpha(d\sigma_i) \int_0^1 ds \frac{d}{ds} \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i),$$

with

$$\frac{d}{ds} \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) = - \sum_{j=1}^l \sum_{u=1}^l F_{ju}(m(\pi_1 \tilde{\nu}_s^*)) \int \pi_1(\tilde{\nu}_1^* - \tilde{\nu}_2^*)(d\bar{\sigma}_i) g_u(\bar{\sigma}_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) (g_j(\sigma_i) - \vartheta_s(g_j)) \tag{79}$$

and where

$$\vartheta_s(g_j) = \int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) g_j(\sigma_i).$$

Therefore it follows from our previous considerations that

$$|\gamma'_1(f|\nu'_1) - \gamma'_1(f|\nu'_2)| \leq \|\partial^2 F\|_{\max, \infty} \delta(g) \|\tilde{\nu}_1^* - \tilde{\nu}_2^*\| \sum_{j=1}^l \int_0^1 ds \int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) |g_j(\sigma_i) - \vartheta_s(g_j)|. \tag{80}$$

By adding and subtracting  $g_j(a_i)$  from the term  $g_j(\sigma_i) - \vartheta_s(g_j)$  and applying the triangle inequality, we obtain

$$\int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) |g_j(\sigma_i) - \vartheta_s(g_j)| \leq 2 \int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) |g_j(\sigma_i) - g_j(a_i)|. \tag{81}$$

Now it follows from Hölder's inequality and the facts that (1) the square root function is concave, (2)  $\hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) \leq e^{\delta_{F,s} k}(\sigma_i, \eta_i)$ , and (3) the Lipschitz property of  $g_j$  that

$$\begin{aligned} \sum_{j=1}^l \int_0^1 ds \int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) |g_j(\sigma_i) - \vartheta_s(g_j)| &\leq 2 \int_0^1 ds \inf_{a_i \in S} \sum_{j=1}^l \left( \int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) \right. \\ &\quad \left. \times (g_j(\sigma_i) - g_j(a_i))^2 \right)^{1/2} \leq 2 \int_0^1 ds \inf_{a_i \in S} \left( \int \alpha'(d\eta_i) \alpha(d\sigma_i) \hat{h}_{\tilde{\nu}_s^*}(\sigma_i, \eta_i) \sum_{j=1}^l (g_j(\sigma_i) - g_j(a_i))^2 \right)^{1/2} \\ &\leq 2 \|g\|_{d,2} e^{\delta_{F,s} l/2} \inf_{a_i \in S} \left( \int \alpha(d\sigma_i) d^2(\sigma_i, a_i) \right)^{1/2}. \end{aligned} \tag{82}$$

Finally putting this bound in (82) into (80) and noting from Proposition 4.2 that  $\|\tilde{\nu}_1^* - \tilde{\nu}_2^*\| \leq L_1 \|\nu'_1 - \nu'_2\|$  follow the proof.  $\square$

**V. EXAMPLES**

We now present two examples for the class of models discussed in the above section. In the first example we consider specific forms of the functions  $F$  and  $g$  and a specific form of the joint *a priori* measure  $K$ . The second example is about general forms of  $F$  and  $g$  and a specific form of the joint *a priori* measure  $K$ .

**A. Short-time Gibbsianness of rotator mean-field models under diffusive time evolution**

The first example we consider is the Curie–Weiss rotator model under sitewise independent diffusive time evolution. Here the single-site spin spaces for both the initial and the transformed systems are the same, i.e.,  $S = S' = S^{q-1}$ , where  $S^{q-1}$  is the sphere in the  $q$ -dimensional Euclidean space with  $q \geq 2$ . The interaction for the initial system is given by

$$\Phi(\nu) = F(\nu[\sigma_i^1], \dots, \nu[\sigma_i^q]) = - \frac{\beta \sum_{j=1}^q \nu[\sigma_i^j]^2}{2}, \tag{83}$$

where  $g_j(\sigma_i) = \sigma_i^j$  is the  $j$ th coordinate of the point  $\sigma_i \in S^{q-1}$  and  $l = q$ .

Next let  $K$  be given by

$$K(d\sigma_i, d\eta_i) = K_t(d\sigma_i, d\eta_i) = k_t(\sigma_i, \eta_i) \alpha_0(d\sigma_i) \alpha_0(d\eta_i), \tag{84}$$

where  $\alpha_0$  is the equidistribution on  $S^{q-1}$  and  $k_t$  is the heat kernel on the sphere, i.e.,

$$(e^{\Delta t} \varphi)(\eta_i) = \int \alpha_0(d\sigma_i) k_t(\sigma_i, \eta_i) \varphi(\sigma_i), \tag{85}$$

where  $\Delta$  is the Laplace–Beltrami operator on the sphere and  $\varphi$  is any test function.  $k_t$  is also called

the *Gauss–Weierstrass kernel*. For more background on the heat kernel on Riemannian manifolds, see the introduction of Ref. 1.

**Lemma 5.1:** *Assuming the above setup the Lipschitz constant  $L$  for the CFLPK is given by*

$$L = L(q, \beta, t) = 4\sqrt{2}q\beta e^\beta(1 - e^{-(q-1)t})^{1/2}. \tag{86}$$

**Proof of Lemma 5.1:** The proof follows by just evaluating the terms appearing in  $C(F, g)$  and  $\rho_\alpha(k)$ . First observe that for the interaction considered in this subsection  $\|\partial^2 F\|_{\max, \infty} = \beta$ ,  $\delta(g) = 2q$ ,  $\|g\|_{d,2} = 1$ , since  $d(\sigma_i, \bar{\sigma}_i) = \|g(\sigma_i) - g(\bar{\sigma}_i)\|_2$  and  $\delta_{F,g} = 2\beta$ . So these give rise to

$$C(F, g) = C(q, \beta) = 4q\beta e^\beta. \tag{87}$$

To obtain  $\rho_\alpha(k)$  we argue as follows: Denote by  $u(t)$  the  $q$ th component of a diffusion  $\sigma_i(t)$  on the sphere started at  $u(0) = 1$  (in the “north pole”) and  $\mathbb{E}$  the expectation with respect to the corresponding diffusion. By reversibility we now choose  $a_i = \eta_i$ , such that  $a_i$  is the north pole. Then it is not hard to see that

$$d^2(\sigma_i(t), a_i) = 2(1 - u(t)).$$

This gives rise to

$$\int \alpha_0(d\sigma) k_t(\sigma_i(t), a_i) d^2(\sigma_i(t), a_i) = 2(1 - \mathbb{E}u(t)). \tag{88}$$

It follows from the above and the rotation invariance of the diffusion on the sphere that to compute  $\rho_\alpha(k)$  we only need the  $q$ th component of the diffusion which according to the Laplace–Beltrami operator on the sphere has generator of the form

$$(1 - u^2) \left( \frac{d}{du} \right)^2 - (q - 1)u \frac{d}{du}. \tag{89}$$

This generates the equation  $(d/dt)\mathbb{E}u_t = -(q - 1)\mathbb{E}u_t$ . Solving this equation with the initial condition  $u(t=0) = 1$  yields

$$\mathbb{E}u(t) = e^{-(q-1)t}, \tag{90}$$

which leads to the desired expression for  $\rho_\alpha(k)$ , i.e.,

$$\rho_\alpha(k) = \sqrt{2}(1 - e^{-(q-1)t})^{1/2} \tag{91}$$

and therefore

$$L = 4\sqrt{2}q\beta e^\beta(1 - e^{-(q-1)t})^{1/2}.$$

□

**Remark:**

- (1) The smallness of  $L$  for this example emanates from at least two sources, namely, small values of  $t$  and  $\beta$ . That is, if  $\beta$  is small enough the system will be Gibbs at all times. However, if we start with large  $\beta$  then we hope to preserve Gibbsianness at only small values of  $t$ .
- (2) We also have for any arbitrary chosen  $a_i$  that

$$\int_S d^2(\sigma_i, a_i) \alpha(d\sigma_i) = \int 2(1 - \sigma \cdot a) \alpha_\sigma(d\sigma) = 2. \tag{92}$$

Thus we have for this example  $\rho_\alpha = \sqrt{2}$ .

**B. Local approximation and preservation of Gibbsianness**

As our second example we start from an initial compact Polish space  $S$  endowed with a metric  $d$  and an *a priori* measure  $\alpha$ . We consider general  $F$  and  $g$  defining the initial Hamiltonian.

We partition the initial space  $S$  into finitely or countably infinitely many disjoint Borel sets with nonzero  $\alpha$  measure indexed by the elements in  $S'$ , i.e.,

$$S = \bigcup_{\eta_i \in S'} S_{\eta_i}, \quad \text{with } \alpha(S_{\eta_i}) > 0 \quad \text{for all } \eta_i \in S'. \tag{93}$$

We then consider the deterministic map  $T: S \rightarrow S'$ , such that  $T(\sigma_i) = \eta_i$  for all  $\sigma_i \in S_{\eta_i}$ . That is, every point is mapped to the label of the class it belongs to. If we start with a finite initial space, this transformation is the so-called fuzzy map which, when starting from an initial Potts model, was studied in Ref. 20. In the present generality this example was studied in Ref. 24 and we want to see here what the mean-field estimates of the present paper provide. Let us formulate the form of the Lipschitz constant  $L$  for the CFLPK resulting from the local approximations in the following lemma.

**Lemma 5.2:** *Assume the setup above; then the Lipschitz constant  $L$  is given by*

$$L = L(F, g, T) = C(F, g) \sup_{\eta_i \in S'} \alpha(S_{\eta_i})^{-1/2} \inf_{a_i \in S_{\eta_i}} \left( \int \alpha|_{S_{\eta_i}}(d\sigma_i) d^2(\sigma_i, a_i) \right)^{1/2}. \tag{94}$$

The proof of the above lemma follows straight away from the definition of the constant  $L$  and observing that  $\alpha(d\sigma_i)k(\sigma_i, \eta_i) = \alpha|_{S_{\eta_i}}(d\sigma_i) / \alpha(S_{\eta_i})$ .

Once again, the constant  $L$  will be small either if the initial interaction is weak enough or if the local approximation is fine enough. For  $L < 1$ , by the general Theorem 4.3, this implies Gibbsianness and continuity estimates of the form (76).

**ACKNOWLEDGMENTS**

The authors thank Aernout van Enter, Roberto Fernández, Frank Redig, and Wioletta Ruszel for interesting discussions.

**APPENDIX: PROOF OF THEOREM 3.3**

The assertion that  $J_{\nu'}$  attains its infimum on  $M_{\nu'}$ , trivially follows from the lower semicontinuity of  $J_{\nu'}$  and compactness of  $M_{\nu'}$ . Now to proceed with the rest of the proof we take any minimizer  $\tilde{\nu}^*$  of  $J_{\nu'}$ , which has the representation  $\tilde{\nu}^*(d\xi_i) = \nu'(d\eta_i)\alpha(d\sigma_i)f_{\nu'}(\sigma_i|\eta_i)$ . Then it remains to show that  $f_{\nu'}$  (1) is  $\alpha \otimes \nu'$  a.s. strictly positive and (2) takes the form (23)  $\alpha \otimes \nu'$  a.s.

(1) We now proceed to show the almost sure strict positivity of the minimizing conditional  $\alpha$ -density  $f_{\nu'}$ . For each  $\eta_i \in S'$  we set  $A_{\eta_i} = \{\sigma_i \in S : f_{\nu'}(\sigma_i|\eta_i) = 0\}$  and denote  $B$  by the set of  $\eta_i$  for which  $\alpha(A_{\eta_i}) > 0$ . The proof consists in establishing a contradiction that  $\tilde{\nu}^*$  is not a minimizer of  $J_{\nu'}$  whenever  $B$  has a positive  $\nu'$  measure. This is an adaptation of arguments found in Ref. 2 and references therein modified to suit our case.

To be precise let  $b: S \rightarrow \mathbb{R}$  be a strictly positive measurable map with  $b \leq 1$ , and for each  $\eta_i \in B$  we define a bounded measurable map  $g(\cdot|\eta_i): S \rightarrow \mathbb{R}$  by

$$g(\sigma_i|\eta_i) = \frac{(1_{A_{\eta_i}^c} f_{\nu'}(\cdot|\eta_i))(\sigma_i) + (1_{A_{\eta_i}} b)(\sigma_i)}{1 + \int_{A_{\eta_i}} b(\hat{\sigma}_i)\alpha(d\hat{\sigma}_i)}. \tag{A1}$$

Now set  $u_{\eta_i} = 1 + \int_{A_{\eta_i}} b(\hat{\sigma}_i)\alpha(d\hat{\sigma}_i)$  and further define for each  $\eta_i \in S'$

$$p_\varepsilon(\cdot, \eta_i) = \begin{cases} \varepsilon g(\cdot | \eta_i) + (1 - \varepsilon) f_{\nu'}(\cdot | \eta_i) = \frac{\varepsilon}{u_{\eta_i}} 1_{A_{\eta_i}} b + C_{\varepsilon, \eta_i} 1_{A_{\eta_i}^c} f_{\nu'}(\cdot | \eta_i) & \text{if } \eta_i \in B \\ f_{\nu'}(\cdot | \eta_i) & \text{if } \eta_i \in B^c, \end{cases} \tag{A2}$$

where

$$C_{\varepsilon, \eta_i} = \varepsilon \left( \frac{1}{u_{\eta_i}} - 1 \right) + 1$$

and  $\varepsilon \in [0, 1]$ . It is easy to check that  $1 \leq u_{\eta_i} \leq 2$ , which implies that  $C_{\varepsilon, \eta_i}$  as well as  $\log C_{\varepsilon, \eta_i}$  are uniformly bounded. Let us set  $\tilde{\nu}_\varepsilon^*(d\xi_i) = \nu'(d\eta_i) \alpha(\sigma_i) p_\varepsilon(\sigma_i, \eta_i)$  and observe from the above that we can write the relative entropy of  $\tilde{\nu}_\varepsilon^*$  with respect to  $\alpha \otimes \nu'$  as

$$\begin{aligned} S(\tilde{\nu}_\varepsilon^* | \alpha \otimes \nu') &= \int_{B^c} \nu'(d\eta_i) S(\tilde{\nu}_\varepsilon^*(\cdot | \eta_i) | \alpha) + \int_B \nu'(d\eta_i) \left\{ \frac{\varepsilon}{u_{\eta_i}} \int_{A_{\eta_i}} \alpha(d\sigma_i) b(\sigma_i) \log b(\sigma_i) \right. \\ &\quad + \frac{\varepsilon}{u_{\eta_i}} \log \frac{\varepsilon}{u_{\eta_i}} \int_{A_{\eta_i}} \alpha(d\sigma_i) b(\sigma_i) + C_{\varepsilon, \eta_i} \int_{A_{\eta_i}^c} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \log f_{\nu'}(\sigma_i | \eta_i) \\ &\quad \left. + C_{\varepsilon, \eta_i} \log C_{\varepsilon, \eta_i} \int_{A_{\eta_i}^c} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \right\}. \end{aligned} \tag{A3}$$

Now we define a function  $h: [0, 1] \rightarrow \mathbb{R}$  by

$$h(\varepsilon) = J_{\nu'}(\tilde{\nu}_\varepsilon^*) = S(\tilde{\nu}_\varepsilon^* | \alpha \otimes \nu') + \Phi(\pi_1 \tilde{\nu}_\varepsilon^*) - \tilde{\nu}_\varepsilon^*[\log k]. \tag{A4}$$

Lebesgue's dominated convergence theorem and the continuity property of  $\Phi^{(1)}$  that  $h$  is continuously differentiable on  $(0, 1)$  follow from the uniform boundedness of  $C_{\varepsilon, \eta_i}$  and  $\log C_{\varepsilon, \eta_i}$ . Observe that  $h(0) = J_{\nu'}(\tilde{\nu}^*)$  and one would expect  $h$  to be decreasing as  $\varepsilon \downarrow 0$ , i.e.,  $h(\varepsilon) - h(0) > 0$  for  $\varepsilon$  close to zero. However, we will show that the converse of the above holds if  $B$  has positive  $\nu'$  measure. More precisely, differentiating  $h$  we obtain for  $\varepsilon \in (0, 1)$

$$\begin{aligned} h'(\varepsilon) &= \int_B \nu'(d\eta_i) \left\{ \frac{1}{u_{\eta_i}} \log \frac{\varepsilon}{u_{\eta_i}} \int_{A_{\eta_i}} \alpha(d\sigma_i) b(\sigma_i) + \left( \frac{1}{u_{\eta_i}} - 1 \right) \log C_{\varepsilon, \eta_i} \int_{A_{\eta_i}^c} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \right\} \\ &\quad + \int_B \nu'(d\eta_i) \int_{A_{\eta_i}} \alpha(d\sigma_i) (g(\sigma_i | \eta_i) - f_{\nu'}(\sigma_i | \eta_i)) \Phi^{(1)}(\pi_1 \tilde{\nu}_\varepsilon^*, \delta_{\sigma_i}) + C(B, b, f_{\nu'}), \end{aligned} \tag{A5}$$

where  $C(B, b, f_{\nu'})$  is a constant which depends on  $B$ ,  $b$ , and  $f_{\nu'}$  but independent of  $\varepsilon$ . Assuming  $\nu'(B) > 0$  then the limit

$$\lim_{\varepsilon \downarrow 0} h'(\varepsilon) = -\infty$$

since the term

$$\int_B \nu'(d\eta_i) \frac{1}{u_{\eta_i}} \log \frac{\varepsilon}{u_{\eta_i}} \int_{A_{\eta_i}} \alpha(d\sigma_i) b(\sigma_i) \tag{A6}$$

goes to negative infinity while the rest remains bounded. This implies that

$$\lim_{\varepsilon \downarrow 0} \frac{h(\varepsilon) - h(0)}{\varepsilon} = -\infty, \tag{A7}$$

giving rise to a contradiction since  $\tilde{\nu}^*$  is a minimizer, hence the assumption that  $\nu'(B) > 0$  is false. This concludes the proof of almost sure strict positivity of  $f_{\nu'}$ .

(2) Next we prove that the conditional  $\alpha$ -density of any minimizer of  $J_{\nu'}$  must satisfy the constrained mean-field equation (23). Let  $f_{\nu'}(\cdot|\cdot)$  be a conditional  $\alpha$ -density of a minimizer  $\tilde{\nu}^*$  of  $J_{\nu'}$  and set

$$Y_{\eta_i} = \Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i}) - \log k(\sigma_i, \eta_i) + \log f_{\nu'}(\sigma_i|\eta_i) + \log \int \exp(-\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\hat{\sigma}_i})) k(\hat{\sigma}_i, \eta_i) \alpha(\hat{\sigma}_i). \tag{A8}$$

We are now left to show that  $\alpha(Y_{\eta_i} = 0) = 1$  for  $\nu'$  almost all  $\eta_i \in S'$ .

The idea of the proof is again to assume the contrary and arrive at the contradiction that a suitable perturbation of the conditional  $\alpha$ -density  $f_{\nu'}$  would have a lower value of  $J_{\nu'}$ . That is, we assume  $\alpha(Y_{\eta_i} \neq 0) > 0$  on some subset  $B$  of  $S'$  with positive  $\nu'$  measure. In a first step this implies that for any  $\eta_i \in B$  both inequalities  $\alpha(Y_{\eta_i} \geq 0) > 0$  and  $\alpha(Y_{\eta_i} < 0) > 0$  must be the case. Indeed, the assumption that, e.g., the second inequality is not true leads to a contradiction. To see this define for each  $\eta_i \in B$  and  $\delta > 0$

$$A_{\eta_i}^\delta = \{\sigma_i \in S : \psi_{\nu'}(\sigma_i|\eta_i) \geq \delta\}, \tag{A9}$$

with

$$\psi_{\nu'}(\sigma_i|\eta_i) = f_{\nu'}(\sigma_i|\eta_i) - \frac{\exp(-\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i})) k(\sigma_i, \eta_i)}{\int \exp(-\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\hat{\sigma}_i})) k(\hat{\sigma}_i, \eta_i) \alpha(\hat{\sigma}_i)}$$

so that we would have  $\lim_{\delta \downarrow 0} \alpha(A_{\eta_i}^\delta) = 1$ .

Taking the  $\alpha$  integral of  $\psi_{\nu'}(\sigma_i|\eta_i)$  yields then the contradiction

$$0 = \int \alpha(d\sigma_i) \psi_{\nu'}(\sigma_i|\eta_i) \geq \int_{A_{\eta_i}^\delta} \alpha(d\sigma_i) \psi_{\nu'}(\sigma_i|\eta_i) \geq \delta \alpha(A_{\eta_i}^\delta) > 0 \tag{A10}$$

for  $\delta$  sufficiently small. This and the  $\alpha \otimes \nu'$  a.s. strict positivity of the minimizing conditional  $\alpha$ -density  $f_{\nu'}$  imply that  $\alpha(1_{\{Y_{\eta_i} < 0\}} f_{\nu'}(\cdot|\eta_i)) > 0$  and  $\alpha(1_{\{Y_{\eta_i} \geq 0\}} f_{\nu'}(\cdot|\eta_i)) > 0$ . Now we set for each  $\eta_i \in S'$

$$C_{\eta_i} = \frac{\alpha(1_{\{Y_{\eta_i} \geq 0\}} f_{\nu'}(\cdot|\eta_i))}{\alpha(1_{\{Y_{\eta_i} < 0\}} f_{\nu'}(\cdot|\eta_i))} \tag{A11}$$

and define for each positive integer  $n \in \mathbb{N}$  the set

$$B_n := \{\eta_i : C_{\eta_i} \in (n - 1, n]\}. \tag{A12}$$

Observe that the  $(B_n)_{n \geq 1}$  is a partition of the set  $B \subset S'$ , i.e.,  $B = \cup_{n \geq 1} B_n$  and the  $B_n$  are pairwise disjoint Borel subsets of  $S'$ . We now consider a perturbation of the minimizing conditional  $\alpha$ -density  $f_{\nu'}(\cdot|\eta_i)$  whose form will be dependent on the choice of the transformed configuration  $\eta_i$ . More precisely, for each  $\varepsilon \in [0, 1]$  we consider the perturbed conditional  $\alpha$ -density  $p_\varepsilon(\sigma_i|\eta_i)$  of the form

$$p_\varepsilon(\cdot|\eta_i) = 1_{B^c} f_{\nu'}(\cdot|\eta_i) + \sum_{n=1}^\infty 1_{B_n}(\eta_i) f_{\varepsilon,n}(\cdot|\eta_i), \tag{A13}$$

where

$$f_{\varepsilon,n}(\cdot|\eta_i) = \left(1 - \frac{\varepsilon}{n}\right) 1_{\{Y_{\eta_i} \geq 0\}} f_{\nu'}(\cdot|\eta_i) + \left(1 + \frac{\varepsilon}{n} C_{\eta_i}\right) 1_{\{Y_{\eta_i} < 0\}} f_{\nu'}(\cdot|\eta_i).$$

Observe that for each  $\eta_i \in S'$ ,  $p_\varepsilon(\cdot|\eta_i)$  is a probability density with respect to  $\alpha$  and  $p_{\varepsilon=0} = f_{\nu'}$ .

As we did in the proof of part (1) we introduce a function  $\varphi: [0, 1] \rightarrow \mathbb{R}$  and show that  $\varphi$  is decreasing for arguments very close to zero whenever  $\nu'(B) > 0$ . However, this would then imply that  $f_{\nu'}$  is not a conditional  $\alpha$ -density for a minimizer of  $J_{\nu'}$ . Hence for the converse to be true,  $\nu'(B) = 0$ , which will then conclude our proof. To formulate this formally we set  $\tilde{\nu}_\varepsilon^*(d\xi_i) = \nu'(d\eta_i) \alpha(d\sigma_i) p_\varepsilon(\sigma_i|\eta_i)$  and define

$$\varphi(\varepsilon) := J_{\nu'}(\tilde{\nu}_\varepsilon^*) = S(\tilde{\nu}_\varepsilon^*|\alpha \otimes \nu') + \Phi(\tilde{\nu}_\varepsilon^*) - \tilde{\nu}_\varepsilon^*[\log k].$$

Now note that the relative entropy takes the form

$$S(\tilde{\nu}_\varepsilon^*|\alpha \otimes \nu') = \int_{B^c} \nu'(d\eta_i) S(\tilde{\nu}_\varepsilon^*(\cdot|\eta_i)|\alpha) + \sum_{n=1}^\infty \int_{B_n} \nu'(d\eta_i) S(\tilde{\nu}_\varepsilon^*(\cdot|\eta_i)|\alpha). \tag{A14}$$

Note that this is a bounded quantity for each  $\varepsilon \in [0, 1]$  since, in particular,

$$S(\tilde{\nu}_\varepsilon^*|\alpha \otimes \nu') \leq 4S(\tilde{\nu}^*|\alpha \otimes \nu') + 2 \log 2.$$

Then by Lebesgue's bounded convergence theorem and the properties of the interaction  $\Phi$ , it is implied that  $\varphi$  is differentiable in the open interval (0,1). However, by the convexity of  $S$  and once again the properties of  $\Phi$ , one can deduce that

$$\varphi'_+(0) = \lim_{\varepsilon \downarrow 0} \frac{\varphi(\varepsilon) - \varphi(0)}{\varepsilon} = \left. \frac{d}{d\varepsilon} J_{\nu'}(\tilde{\nu}_\varepsilon^*) \right|_{\varepsilon=0}.$$

Thus the rest of the proof then shows that  $\varphi'_+(0)$  has a negative sign whenever  $\nu'(B) > 0$ . Let us now evaluate this quantity. Note that because of the form of the perturbed conditional  $\alpha$ -density, the part of  $S(\tilde{\nu}_\varepsilon^*|\alpha \otimes \nu')$  that will play a role in determining  $\varphi'_+(0)$  is the part involving the  $B_n$ . For each  $n$  we set

$$S_n(\varepsilon) = \int_{B_n} \nu'(d\eta_i) S(\tilde{\nu}_\varepsilon^*(\cdot|\eta_i)|\alpha) \tag{A15}$$

and evaluate

$$\begin{aligned} \left. \frac{d}{d\varepsilon} S_n(\varepsilon) \right|_{\varepsilon=0} &= \frac{1}{n} \left\{ \int_{B_n} \nu'(d\eta_i) C_{\eta_i} \int_{Y_{\eta_i} < 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i|\eta_i) \log f_{\nu'}(\sigma_i|\eta_i) \right. \\ &\quad \left. - \int_{B_n} \nu'(d\eta_i) \int_{Y_{\eta_i} \geq 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i|\eta_i) \log f_{\nu'}(\sigma_i|\eta_i) \right\}. \end{aligned} \tag{A16}$$

Next we also evaluate

$$\begin{aligned} \left. \frac{d}{d\varepsilon}(\Phi(\tilde{\nu}_\varepsilon^*) - \tilde{\nu}_\varepsilon^*[\log k]) \right|_{\varepsilon=0} &= \sum_{n=1}^{\infty} \frac{1}{n} \left\{ - \int_{B_n} \nu'(d\eta_i) \int_{Y_{\eta_i} \geq 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) (\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i}) \right. \\ &\quad - \log k(\sigma_i, \eta_i)) + \int_{B_n} \nu'(d\eta_i) C_{\eta_i} \int_{Y_{\eta_i} < 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) (\Phi^{(1)} \\ &\quad \left. \times (\pi_1 \tilde{\nu}^*, \delta_{\sigma_i}) - \log k(\sigma_i, \eta_i)) \right\}. \end{aligned} \tag{A17}$$

It follows from the above considerations that

$$\begin{aligned} \left. \frac{d}{d\varepsilon}(S(\tilde{\nu}_\varepsilon^* | \alpha \otimes \nu') + \Phi(\tilde{\nu}_\varepsilon^*) - \tilde{\nu}_\varepsilon^*[\log k]) \right|_{\varepsilon=0} &= \sum_{n=1}^{\infty} \frac{1}{n} \left\{ - \int_{B_n} \nu'(d\eta_i) \int_{Y_{\eta_i} \geq 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \right. \\ &\quad \times (\log f_{\nu'}(\sigma_i | \eta_i) + \Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i}) - \log k(\sigma_i, \eta_i)) \\ &\quad + \int_{B_n} \nu'(d\eta_i) C_{\eta_i} \int_{Y_{\eta_i} < 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \\ &\quad \left. \times (\log f_{\nu'}(\sigma_i | \eta_i) + \Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i}) - \log k(\sigma_i, \eta_i)) \right\}. \end{aligned} \tag{A18}$$

Furthermore, set

$$r(\eta_i) = \log \int \alpha(d\sigma_i) \exp(-\Phi^{(1)}(\pi_1 \tilde{\nu}^*, \delta_{\sigma_i})) k(\sigma_i, \eta_i).$$

Finally, it follows from (A8) and (A11) after adding and subtracting  $r(\eta_i)$  from the integrands of the  $\alpha$ -integrals in (A18) that

$$\begin{aligned} \left. \frac{d}{d\varepsilon}(S(\tilde{\nu}_\varepsilon^* | \alpha \otimes \nu') + \Phi(\tilde{\nu}_\varepsilon^*) - \tilde{\nu}_\varepsilon^*[\log k]) \right|_{\varepsilon=0} &= \sum_{n=1}^{\infty} \frac{1}{n} \left\{ \int_{B_n} \nu'(d\eta_i) r(\eta_i) \int_{Y_{\eta_i} \geq 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \right. \\ &\quad - \int_{B_n} \nu'(d\eta_i) \int_{Y_{\eta_i} \geq 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) Y_{\eta_i} \\ &\quad - \int_{B_n} \nu'(d\eta_i) C_{\eta_i} r(\eta_i) \int_{Y_{\eta_i} < 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) \\ &\quad \left. + \int_{B_n} \nu'(d\eta_i) C_{\eta_i} \int_{Y_{\eta_i} < 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) Y_{\eta_i} \right\} \\ &= \sum_{n=1}^{\infty} \frac{1}{n} \left\{ - \int_{B_n} \nu'(d\eta_i) \int_{Y_{\eta_i} \geq 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) Y_{\eta_i} \right. \\ &\quad \left. + \int_{B_n} \nu'(d\eta_i) C_{\eta_i} \int_{Y_{\eta_i} < 0} \alpha(d\sigma_i) f_{\nu'}(\sigma_i | \eta_i) Y_{\eta_i} \right\} < 0. \end{aligned} \tag{A19}$$

Therefore we have succeeded in showing that

$$\varphi'_+(0) = \lim_{\varepsilon \downarrow 0} \frac{\varphi(\varepsilon) - \varphi(0)}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{J_{\nu'}(\tilde{\nu}_\varepsilon^*) - J_{\nu'}(\tilde{\nu}^*)}{\varepsilon} < 0$$

whenever  $\nu'(B) > 0$ , contradicting the initial claim that  $\tilde{\nu}^*$  is a minimizer of  $J_{\nu'}$ . Hence for  $\tilde{\nu}^*$  to be a minimizer,  $\nu'(B) = 0$ . This concludes the proof of the claim that the conditional  $\alpha$ -densities  $f_{\nu'}$  of the minimizers of  $J_{\nu'}$  satisfy the constrained mean-field equation (23)  $\alpha \otimes \nu'$  a.s.  $\square$

- <sup>1</sup>Auscher, P., Coulhon, T., Duong, X. T., and Hofmann, S., "Riesz transform on manifolds and heat Kernel regularity," *Ann. Sci. Ec. Normale Super.* **37**, 911 (2004).
- <sup>2</sup>van den Berg, M., Dolars, T. C., Lewis, J. T., and Pulè, J. V., "A perturbed mean field model of an interacting boson gas and the large deviation principle," *Commun. Math. Phys.* **127**, 41 (1990).
- <sup>3</sup>Comets, F., "Large deviation estimates for a conditional probability distribution: Applications to random interaction Gibbs measures," *Probab. Theory Relat. Fields* **80**, 407 (1989).
- <sup>4</sup>Dereudre, D. and Roelly, S., "Propagation of Gibbsianness for infinite-dimensional gradient Brownian diffusions," *J. Stat. Phys.* **121**, 511 (2005).
- <sup>5</sup>Deuschel, J.-D., "Infinite-dimensional diffusion processes as Gibbs measures on  $C[0, 1]^{\mathbb{Z}^d}$ ," *Probab. Theory Relat. Fields* **76**, 325 (1987).
- <sup>6</sup>Dobruschin, R. L., "The description of a random field by means of conditional probabilities and conditions of its regularity," *Theor. Probab. Appl.* **13**, 197 (1968).
- <sup>7</sup>Ellis, R. S., *Entropy, Large Deviations, and Statistical Mechanics*, 1st ed. (Springer-Verlag, New York, 1985); *Entropy, Large Deviations, and Statistical Mechanics* (Springer-Verlag, New York, 2006), Vol. XVIII.
- <sup>8</sup>van Enter, A. C. D., Fernández, R., and Sokal, A. D., "Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory," *J. Stat. Phys.* **72**, 879 (1993).
- <sup>9</sup>van Enter, A. C. D., Fernández, R., den Hollander, F., and Redig, F., "Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures," *Commun. Math. Phys.* **226**, 101 (2002).
- <sup>10</sup>van Enter, A. C. D. and Külske, C., "Two connections between random systems and non-Gibbsian measures," *J. Stat. Phys.* **126**, 1007 (2007).
- <sup>11</sup>van Enter, A. C. D. and Ruszel, W. M., e-print arXiv:0711.3621; "Gibbsianness vs. non-Gibbsianness of time-evolved planar rotor models," *Stochastic Proc. Appl.* (to be published).
- <sup>12</sup>van Enter, A. C. D., Redig, F., and Verbitskiy, E., e-print arXiv:0804.4060; "Gibbsian and non-Gibbsian states at Eurandom," *Statistica Neerlandica* (to be published).
- <sup>13</sup>Fannes, M., Spohn, H., and Verbeure, A., "Equilibrium states for mean field models," *J. Math. Phys.* **21**, 355 (1980).
- <sup>14</sup>Fernández, R., Les Houches LXXXIII, 2005 (unpublished).
- <sup>15</sup>Fernández, R. and Maillard, G., "Construction of a specification from its singleton part," *ALEA Lat. Am. J. Probab. Math. Stat.* **2**, 297 (2006).
- <sup>16</sup>Fröhlich, J., Simon, B., and Spencer, T., "Infrared bounds, phase transitions and continuous symmetry breaking," *Commun. Math. Phys.* **50**, 79 (1976).
- <sup>17</sup>Georgii, H.-O., *Gibbs Measures and Phase Transitions*, de Gruyter Studies in Mathematics Vol. 9 (Walter de Gruyter, Berlin, 1988).
- <sup>18</sup>Griffiths, R. B. and Pearce, P. A., "Position-space renormalization-group transformations: Some proofs and some problems," *Phys. Rev. Lett.* **41**, 917 (1978).
- <sup>19</sup>Griffiths, R. B. and Pearce, P. A., "Mathematical properties of position-space renormalization-group transformations," *J. Stat. Phys.* **20**, 499 (1979).
- <sup>20</sup>Häggström, O. and Külske, C., "Gibbs properties of the fuzzy Potts model on trees and in mean field," *Markov Processes Relat. Fields* **10**, 477 (2004).
- <sup>21</sup>Israel, R. B., *Convexity in the Theory of Lattice Gases*, Princeton Series in Physics (Princeton University Press, Princeton, NJ, 1979).
- <sup>22</sup>Külske, C., "Non-Gibbsianness and phase transition in random lattice spin models," *Markov Processes Relat. Fields* **5**, 357 (1999).
- <sup>23</sup>Külske, C., "Analogues of non-Gibbsianness in joint measures of disordered mean field models," *J. Stat. Phys.* **112**, 1079 (2003).
- <sup>24</sup>Külske, C. and Opoku, A. A., "The posterior metric and the goodness of Gibbsianness for transforms of Gibbs measures," e-print arXiv:0711.3764 ; *Electron. J. Probab.* **13**, 1307 (2008).
- <sup>25</sup>Külske, C. and Le Ny, A., "Spin-flip dynamics of the Curie-Weiss model: Loss of Gibbsianness with possibly broken symmetry," *Commun. Math. Phys.* **271**, 431 (2007).
- <sup>26</sup>Külske, C. and Redig, F., "Loss without recovery of Gibbsianness during diffusion of continuous spins," *Probab. Theory Relat. Fields* **135**, 428 (2006).
- <sup>27</sup>Le Ny, A. and Redig, F., "Short-time conservation of Gibbsianness under local stochastic evolution," *J. Stat. Phys.* **109**, 1073 (2002).
- <sup>28</sup>Parthasarathy, K. R., *Probability Measures on Metric Spaces* (Academic, New York, 1967).
- <sup>29</sup>Rudin, W., *Functional Analysis*, 2nd ed. (McGraw-Hill, New York, 1991).