JOURNAL OF MATHEMATICAL PHYSICS **58**, 023301 (2017)

Lattice gas models with long range interactions

David Aristoff[1,a)] and Lingjiong Zhu[2,b)]
[1]*Department of Mathematics, Colorado State University, 1874 Campus Delivery, Fort Collins, Colorado 80523, USA*
[2]*Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, Florida 32306, USA*

(Received 27 May 2016; accepted 19 January 2017; published online 6 February 2017)

We study microcanonical lattice gas models with long range interactions, including power law interactions. We rigorously obtain a variational principle for the entropy. In a one dimensional example, we find a first order phase transition by proving the entropy is non-differentiable along a certain curve. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4975338]

I. INTRODUCTION

In this article we study lattice gas models with certain long range pair interactions. Our models are generalizations of certain mean field and random graph models, in which all sites or nodes interact with all others with equal strength. In contrast with mean field models, we allow the interaction strength to decay, but at a rate sufficiently slow that interactions between far away sites are still significant.

Our models can be described as follows. Consider the cubic lattice in d dimensions, rescaled so that the spacing between adjacent sites is $1/n$. We consider configurations of particles on the $\sim n^d$ lattice sites that fit inside the d-dimensional unit cube $C = [0,1]^d$. Each configuration consists of an arrangement of particles on the lattice sites in C. We say a site is occupied if there is a particle there; each lattice site is either occupied by one particle or unoccupied. Configurations are assigned an energy from a pair potential $\psi : [0, \sqrt{d}] \to \mathbb{R}$. Interactions are only between occupied sites. The energy associated with distinct occupied sites x and y is $\psi(|x-y|)$, and the total energy of a configuration is obtained by summing $\psi(|x-y|)$ over all occupied sites x, y. We assume that there is $0 < r < d$ such that $\psi(t)$ diverges at least as slowly as t^{-r} as $t \to 0$. As a consequence, the interactions between sites that are far apart (relative to the lattice spacing) make nontrivial contributions to the total energy. Such interactions are sometimes called *long range*.[4,8,12,15] Equivalence of thermodynamic ensembles breaks down in this regime,[3,5] so we will consider only the microcanonical ensemble, in which energy density and site occupancy density are held fixed.

When ψ is a constant function, our model is a microcanonical version of the Curie-Weiss mean field Ising model. When $d = 2$, our model is related to certain random graphs. This is because the occupancy pattern on the lattice can be mapped to an adjacency matrix for a graph on n nodes, where an occupied site is an edge in the graph, and an unoccupied site corresponds to the absence of an edge. In this case, ψ corresponds to an interaction between graph edges. For discussion on the connection to the Curie-Weiss model and random graphs, see the comments in Section V.

We now give a rough description of our main results. Each configuration can be described by an occupancy pattern — a function with value 1 at each occupied site and 0 at each unoccupied site. More precisely, each configuration corresponds to an occupancy pattern f defined on C as follows: if site x is occupied (respectively, unoccupied), f is equal to 1 (respectively, 0) on a d-dimensional cube of side length $1/n$ centered at x. The associated particle density is $\int_C f(x)\,dx$ and the energy density can be approximated by

[a)]E-mail address: aristoff@math.colostate.edu
[b)]E-mail address: zhu@math.fsu.edu

$$n^d \int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy.$$

The potential ψ is divided by n^d so that the energy density scales appropriately as $n \to \infty$. In this limit, there is a continuum of lattice sites in C, and f becomes an occupancy density function with values in $[0,1]$. One can imagine that such f is obtained by smoothing out occupancy pattern functions for large finite n, with $f(x)$ representing the probability that site x is occupied. The entropy density associated with the occupancy density f is

$$-\int_C [f(x)\log f(x) + (1-f(x))\log(1-f(x))]\,dx. \tag{1.1}$$

The entropy density is simply the log of the number of configurations with density approximately given by f, normalized by n^d. The above heuristics show that as $n \to \infty$, the energy density, entropy density, and particle density can be accurately expressed in terms of the occupancy density $f : C \to \mathbb{R}$. If these heuristics are correct, one expects that as $n \to \infty$, at fixed energy and particle density, configurations will have occupancy densities that approach optimizers f_* of the entropy density (1.1) subject to the constraints on particle and energy density. It is then straightforward to write the corresponding Euler-Lagrange equations and, in principle at least, find the optimizers f_*. For details on the above ideas from the point of view of large deviations theory, see for instance Ref. 4.

The main contribution of this paper is to show that the above heuristics can be made mathematically rigorous under very weak assumptions on the potential energy ψ. (Though we focus on lattice gas models, our arguments easily adapt to other long range interacting models, *e.g.,* the α-Ising model.[3] See also Ref. 23 for a closely related result.) We also give an example of an interaction for which the model has a first order phase transition. To our knowledge, such transitions had not yet been rigorously demonstrated in microcanonical models of this type.

Rigorous results in the long range setting described above are relatively scarce. We mention that similar rigorous results have been proved in microcanonical spin models[23] and in the grand canonical Ising model with Kac interactions.[6,7,32] See also Refs. 3 and 11 for similar work on the α-Ising model, and Refs. 24 and 25 for studies of more general long range interacting Ising models. For rigorous analysis of other mean field type models, see Refs. 7, 9, 10, and 17. We also mention related work on random graph models in which the interaction depends on the number of edges and other subgraphs; see Refs. 2, 13, 18, 19, and 26–28.

This article is organized as follows. In Section II we describe our models in detail. We present a variational principle for the entropy and the corresponding Euler-Lagrange equations in Section III. Using these results, we show a phase transition in a one dimensional model in Section IV. In Section V, we discuss connections between our models and certain mean field and random graph models. All proofs are in Section VI.

II. NOTATION AND ASSUMPTIONS

Fix a dimension $d \geq 1$, and for $n \geq 1$ define the lattice $\Lambda_n = \{1,\ldots,n\}^d$. Lattice sites (that is, elements of Λ_n) will be denoted by I, J. Each lattice site I can be occupied or not. A particle configuration is an assignment of occupancy to each site. More precisely, a particle configuration is a function $\eta : \Lambda_n \to \{0,1\}$. (We sometimes write η_n to emphasize the dependence on n.) Here, $\eta(I) = 1$ if site I is occupied, and $\eta(I) = 0$ otherwise. Recall $\psi : [0,\sqrt{d}] \to \mathbb{R}$ is a given pair potential. The interaction between sites $I, J \in \Lambda_n$ is defined by

$$\phi_n(I,J) := \psi\left(n^{-1}|I-J|\right), \tag{2.1}$$

where $|\cdot|$ is the usual Euclidean norm in $\mathbb{R}^d$. Recall $C = [0,1]^d$ is the d-dimensional unit cube. Let C_I be a d-dimensional cube of side length $1/n$ centered (approximately) at $n^{-1}I$. More precisely, $C_I = \{x \in C : I - 1 \leq nx < I\}$, where 1 represents the all ones vector and the inequalities are componentwise. Throughout, we will associate a particle configuration η to a occupancy density function f^η obtained by setting f^η equal to 1 on cubes C_I corresponding to occupied sites I, and 0

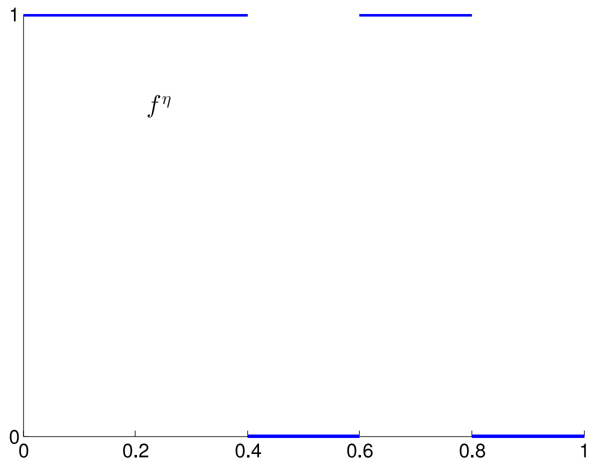

FIG. 1. The graph of the function f^η when $d = 1$, $n = 5$, $\eta(1) = \eta(2) = \eta(4) = 1$, and $\eta(3) = \eta(5) = 0$.

otherwise. More precisely, $f^\eta : C \to [0,1]$ is defined by

$$f^\eta(x) = \eta(I), \qquad \text{if } x \in C_I.$$

See Figure 1. Let $\mathbb{P}_n$ be the uniform probability measure on particle configurations,

$$\mathbb{P}_n(\eta) = 2^{-n^d}, \qquad \text{for all } \eta : \Lambda_n \to \{0,1\}.$$

This defines an equivalent measure on occupancy density functions. That is, under the map $\eta \to f^\eta$, $\mathbb{P}_n$ pushes forward to a probability measure on the space of measurable functions $C \to [0,1]$. We denote this measure by the same symbol, $\mathbb{P}_n$, since no confusion should arise. Define the energy density E_n of $\eta : \Lambda_n \to \{0,1\}$ as the sum of $\phi_n(I,J)$ over all pairs of occupied sites I, J, appropriately normalized,

$$E_n(\eta) = n^{-2d} \sum_{I,J \in \Lambda_n} \eta(I)\eta(J)\phi_n(I,J). \tag{2.2}$$

Define also the particle density N_n of $\eta : \Lambda_n \to \{0,1\}$ as the average site occupancy,

$$N_n(\eta) = n^{-d} \sum_{I \in \Lambda_n} \eta(I). \tag{2.3}$$

Fix parameters $\xi, \rho \in \mathbb{R}$, and define the microcanonical entropy

$$S(\xi,\rho) = \lim_{\delta \to 0^+} \lim_{n \to \infty} n^{-d} \log \mathbb{P}_n \left(E_n(\eta) \in (\xi - \delta, \xi + \delta),\, N_n(\eta) \in (\rho - \delta, \rho + \delta)\right). \tag{2.4}$$

We show below the limit defining $S(\xi,\rho)$ exists under the following assumption.

Assumption 1. The map $(x,y) \mapsto \psi(|x-y|)$ is in $L^q(C^2)$ for some $q > 1$. Moreover, it is Riemann integrable.

Assumption 1 will hold throughout the remainder of the paper. As a typical example of interactions satisfying this assumption, we keep in mind the case of power law interactions in dimension $d = 1$, where $\psi(t) = t^{-r}$ for $t \in (0,1]$ and $\psi(0) = 0$, with $r \in (0,1)$ constant. (We set $\psi(0) = 0$ simply so that a particle does not interact with itself.) We note, however, that the interaction need not even be continuous. In Section IV, we consider a modified power law interaction for which the entropy S is singular.

Before proceeding we comment on boundary conditions. Note that the definitions (2.2) and (2.3) suggests free boundary conditions. However, we note that periodic boundary conditions can be obtained by replacing Euclidean distance $|\cdot|$ in (2.1) with distance on the flat torus $\mathbb{R}^d/\mathbb{Z}^d$. In dimension $d = 1$, this corresponds to setting $\psi(t) = \psi(1-t)$ for all $t \in [0,1]$. See Ref. 3 for a similar discussion of boundary conditions in the α-Ising model.

III. LARGE DEVIATIONS, ENTROPY, AND EULER-LAGRANGE EQUATIONS

Before stating our results we give a sketch of the arguments. First, we show that the formula (1.1) represents the logarithm of the number of configurations with occupancy density approximately equal to f. The relevant result is Theorem 2. Roughly, for large n and suitable collections A of particle configurations, we show that

$$n^{-d}\log \mathbb{P}_n(f^\eta \in A) \approx \sup_{f\in A}\left\{-\log 2 - \int_C [f(x)\log f(x) + (1-f(x))\log(1-f(x))]\,dx\right\}. \tag{3.1}$$

The extra term log 2 comes from the probability normalization. Equation (3.1) follows from a *large deviations principle*, and the quantity in brackets (multiplied by −1) is called the *rate function*. See below for precise definitions of this terminology.

The trick is to prove the large deviations principle in a topology strong enough so that the set of suitable collections contains the collection A we are interested in. Since we want to compute the microcanonical entropy (2.4), we take A to be the collection of configurations with energy density and particle density approximately equal to ξ and ρ, respectively. If the (approximate) energy density $\int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy$ and particle density $\int_C f(x)\,dx$ are continuous in f, then A is indeed suitable and we can use (3.1) to compute the microcanonical entropy (2.4). This is a consequence of the well-known *contraction principle*[16,31] in large deviations theory. (We state the contraction principle in Section VI.) It turns out that particle density is continuous in any reasonable topology, but it is not trivial to show energy density is continuous in a suitable topology. See Lemma 11. There are some additional technical issues associated with showing the energy density can be well approximated by $\int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy$ in the sense of exponential equivalence[31] (this term is defined in Section VI). See Lemma 12.

We are now ready to state our results. For notational convenience, we write

$$H_{\text{bin}}(t) = \begin{cases} t\log t + (1-t)\log(1-t) + \log 2, & t\in[0,1] \\ \infty, & t\notin[0,1] \end{cases}.$$

(We write H_{bin} here because of the similarity to the binary entropy function $-t\log_2 t - (1-t)\log_2(1-t)$.) Before proceeding, we introduce the terminology we need from large deviations theory. A sequence $\mathbb{Q}_n$ of probability measures on a topological space $\mathcal{T}$ is said to satisfy a large deviations principle with speed a_n and rate function $K:\mathcal{T}\to\mathbb{R}$ if K is non-negative and lower semicontinuous, and for any measurable set $A\subset\mathcal{T}$,

$$-\inf_{x\in A^\circ} K(x) \le \liminf_{n\to\infty} a_n^{-1}\log\mathbb{Q}_n(A) \le \limsup_{n\to\infty} a_n^{-1}\log\mathbb{Q}_n(A) \le -\inf_{x\in\bar{A}} K(x), \tag{3.2}$$

where A° denotes the interior of A and $\bar{A}$ the closure of A. We refer to the first inequality in (3.2) as the lower bound and the last inequality in (3.2) as the upper bound. Note that compared to the description above, we have replaced $\sup -K$ with $-\inf K$. This is so that we can be consistent with standard notations in large deviations theory.

Throughout we fix $p\in[1,\infty)$ and $q\in(1,\infty]$ with $p^{-1}+q^{-1}=1$. We prove a large deviations principle for $\mathbb{P}_n$ on the Banach space of functions in $L^p(C)$ endowed with the weak topology. We denote this space by $\mathcal{X}$. We will also consider the subset $\mathcal{Y}=\{f\in\mathcal{X} : f(x)\in[0,1] \text{ for a.e. } x\}\subset\mathcal{X}$. Unless otherwise specified, we endow $\mathcal{Y}$ with the subspace topology.

Theorem 2. *The sequence $\mathbb{P}_n$ satisfies a large deviations principle on $\mathcal{X}$ with speed n^d and rate function*

$$H(f) = \int_C H_{\text{bin}}(f(x))\,dx.$$

Consider the following constrained subset of $\mathcal{Y}$:

$$\mathcal{Y}_{\xi,\rho} := \left\{f\in\mathcal{Y} : \int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy = \xi, \int_C f(x)\,dx = \rho\right\}.$$

Abusing notation, we refer to $\int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy$ as the energy density, even though when $f = f^\eta$ this expression is not exactly equal to (2.2). We show in Lemma 12 that they are nonetheless close in the sense of exponential equivalence (this term is defined precisely above Lemma 9 in Section VI). Note that $\int_C f^\eta(x)\,dx$ is exactly equal to the particle density of η defined in (2.3). Thus, we think of $\mathcal{Y}_{\xi,\rho}$ as the collection of occupancy density functions f with energy density ξ and particle density ρ. As discussed above, exponential equivalence, Theorem 2, and the contraction principle lead to the following variational expression for the entropy.

Theorem 3. *We have*

$$S(\xi,\rho) = -\inf_{f\in\mathcal{Y}_{\xi,\rho}} H(f) = \sup_{f\in\mathcal{Y}_{\xi,\rho}} [-H(f)], \tag{3.3}$$

with the infimum over the empty set equal to ∞ *by convention.*

We note that a very similar rigorous result was recently proved, using direct arguments, in Ref. 23. In our proof, we use the machinery of large deviations theory, proving a large deviations principle for $\mathbb{P}_n$ and using the contraction principle and exponential equivalence to get a variational principle for the entropy. Compared to the result in Ref. 23, our assumptions on the interaction ψ are weaker; in particular, we allow for interactions ψ that are non-smooth away from 0. On the other hand, the article[23] considers different domain shapes as well as more general short range interactions. While many of our arguments could be generalized in this way, we do not pursue this direction, partly because of our interest in the connection of our problem with random graph models (in which a square domain represents an adjacency matrix).

Below we will refer to functions $f_* \in \mathcal{Y}_{\xi,\rho}$ with $S(\xi,\rho) = -H(f_*)$ as optimizers of the variational problem (3.3). Optimizers represent the most likely structure of large particle configurations. For instance, if f_* is the unique optimizer of (3.3) and n is large, then $f_*(n^{-1}I)$ is roughly the probability that $\nu(I) = 1$, *i.e.*, there is a particle at site $I \in \Lambda_n$.

Standard results in the calculus of variations lead to the following. Whenever (ξ,ρ) corresponds to achievable values of energy and particle density, compactness arguments show that optimizers of the variational problem (3.3) exist; moreover optimizers in the interior of the appropriate function space satisfy the Euler-Lagrange equations. To make these statements precise, we define

$$\Omega = \left\{(\xi,\rho) \,:\, \mathcal{Y}_{\xi,\rho} \neq \emptyset\right\},$$

as the region of achievable energy and particle densities, and write

$$\mathcal{F} = \{f \in \mathcal{Y} \,:\, \exists\epsilon > 0 \text{ s.t. } f(x) \in [\epsilon, 1-\epsilon] \text{ for a.e. } x\}$$

for the interior of $\mathcal{Y}$ with respect to the essential sup norm.

Theorem 4. *Optimizers of (3.3) exist whenever* $(\xi,\rho) \in \Omega$. *If* $f_* \in \mathcal{F} \cap \mathcal{Y}_{\xi,\rho}$ *is an optimizer, then for a.e. x, either*

$$f_*(x) = \frac{\exp\left(\mu + \beta\int_C f_*(y)\psi(|x-y|)\,dy\right)}{1+\exp\left(\mu + \beta\int_C f_*(y)\psi(|x-y|)\,dy\right)}, \tag{3.4}$$

for some $\beta,\mu \in \mathbb{R}$ *with* $(\beta,\mu) \neq (0,0)$, *or*

$$\int_C f_*(y)\psi(|x-y|)\,dy \equiv \xi/\rho. \tag{3.5}$$

Viewing the expression in (3.4) as a convolution leads to the following corollary.

Corollary 5. If the Euler-Lagrange equation (3.4) holds, then f_* *is continuous.*

Intuitively, Equation (3.5) holds only when the constraints take over in the optimization (see the discussion following Theorem 9.1 in Ref. 14), so in "most" cases we expect that instead the Euler-Lagrange equations (3.4) hold. Corollary 5 has the following interesting consequence. Suppose that $(\xi,\rho) \in \Omega$ and $\xi \neq \lambda\rho^2$, where $\lambda := \int_{C^2}\psi(|x-y|)\,dx\,dy$. Then the constant valued

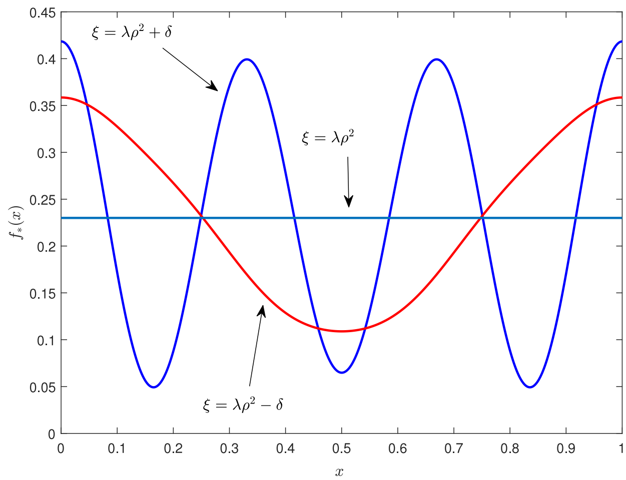

FIG. 2. Optimizers f_* of $S(\xi,\rho)$ (computed numerically) when ψ is given by Assumption 6 with $r = 1/2$ and $M = 10$. The plots show optimizers f_* of $S(\xi,\rho)$ at fixed $\rho = 0.23$ and 3 different values of ξ: on the transition curve ($\xi = \lambda\rho^2$) as well as just below ($\xi = \lambda\rho^2 - \delta$) and just above ($\xi = \lambda\rho^2 + \delta$) the transition curve (here $\delta = 0.02$).

occupation density $f \equiv \rho$ cannot be an optimizer, since it has energy density

$$\int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy = \rho^2 \int_{C^2} \psi(|x-y|)\,dx\,dy = \lambda\rho^2 \neq \xi.$$

Thus, when $\xi \neq \lambda\rho^2$, any optimizer f_* of $S(\xi,\rho)$ must be nonconstant. Suppose such f_* satisfies the Euler-Lagrange equations (3.4). Then f_* is continuous and nonconstant, say with two distinct values $a < b$, so Corollary 5 and the intermediate value theorem show that f_* takes every value in the interval $[a,b]$. In particular, f_* cannot be constant or piecewise constant. Such optimizers f_* have a spatially inhomogeneous occupation density profile. (See Ref. 3 and Figure 2 for examples where the optimizer has a curved structure.) Note the contrast with typical short range interactions, for which optimizers of the entropy have a spatially homogeneous (in pure phases) or piecewise homogenous (in mixed phases) density profile as the system size goes to infinity.

It is also interesting to consider the case of constant valued interactions. Suppose $\psi \equiv \lambda$ is constant. Then one of the constraints is redundant: if $\int_C f(x)\,dx = \rho$, then

$$\int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy = \lambda\rho^2.$$

Thus, particle density ρ completely determines energy density ξ via $\xi = \lambda\rho^2$. In this case, it is easy to see that the only optimizer of the entropy $S(\lambda\rho^2,\rho)$ is the constant function $f_* \equiv \rho$.

IV. SINGULARITY OF THE ENTROPY IN A ONE DIMENSIONAL EXAMPLE

Here we consider an example in dimension $d = 1$ in which the entropy S is singular. We will consider ψ with the following modified power law structure.

Assumption 6. For some constants $r \in (0,1)$ and $M > 0$,

$$\psi(t) = \begin{cases} t^{-r}, & 0 < t < 1/4 \\ M, & 1/4 \le t \le 1/2 \end{cases},$$

and $\psi(0) = 0$. Also, ψ is symmetric: for each $t \in [0,1]$, $\psi(t) = \psi(1-t)$.

Note that symmetry of ψ corresponds to periodic boundary conditions for the particle configurations, *i.e.*, particle configurations on a circle. Clearly, ψ satisfies Assumption 1. If M is chosen carefully, then at a given particle density, at high energy configurations tend to be multimodal, while at low energy configurations tend to be unimodal; see Figure 2. The switch from unimodal to multimodal structure corresponds to a singularity in the entropy, as we show in Theorem 8. To make this argument rigorous, we need two ingredients. First, we identify where the interface between unimodal and multimodal structure should occur. The simplest guess is that the interface corresponds to parameter values (ξ,ρ) at which the optimizers are constant valued occupation densities $f \equiv \rho$. This guess turns out to be correct, as we show below. And second, we have to verify that parameters (ξ,ρ) on both sides of this interface are achievable, so that the transition interface is in the interior of Ω. We prove this in Lemma 7.

Before proceeding with the proof we introduce some notation. We write

$$\lambda = \int_{[0,1]^2} \psi(|x-y|)\,dx\,dy$$

for the integrated interaction function, and we define

$$\xi(f) = \int_{[0,1]^2} f(x)f(y)\psi(|x-y|)\,dx\,dy.$$

When $\xi = \lambda\rho^2$, the constant function $f \equiv \rho$ satisfies the constraints and is therefore an optimizer of the entropy. The curve $\xi = \lambda\rho^2$ is the interface between unimodal and multimodal optimizers discussed above, and it corresponds to a singularity in the microcanonical entropy, as we show in Theorem 8. We now show this interface lies in the interior of Ω, at least for a range of densities ρ.

Lemma 7. For each $r \in (0,1)$, *there is an interaction* ψ *satisfying Assumption 6 with the following property. There is* $\epsilon > 0$ *such that the curve*

$$\{(\xi,\rho) \in \Omega : \xi = \lambda\rho^2, \rho \in (1/4-\epsilon, 1/4)\}$$

is in the interior of Ω.

Lemma 7 is proved by exhibiting functions f which integrate to ρ and have values of $\xi(f)$ both larger and smaller than $\xi(\rho) = \lambda\rho^2$. Such functions can be found for suitable M. We do not attempt to find the complete interior or boundary of Ω. Fortunately, Lemma 7 suffices for the following.

Theorem 8. *Let* ψ *be as in Lemma 7 with* $r < 1/2$. *Then the entropy S is non-differentiable along the curve* $\{(\xi,\rho) \in \Omega : \xi = \lambda\rho^2, \rho \in (1/4-\epsilon, 1/4)\}$.

Note that we needed Lemma 7 to show that the curve $\xi = \lambda\rho^2$ is in the interior of Ω for $\rho \in (1/4-\epsilon, 1/4)$; otherwise, the notion of a singularity along $\xi = \lambda\rho^2$ does not make sense. Theorem 8 shows there is a first order phase transition, *i.e.*, a discontinuity in the first derivative of the entropy, across this curve. The curve corresponds to optimizers that are constant valued. (Recall from Corollary 5 that optimizers must be non-constant off this curve.) The first order transition corresponds to a qualitative change in the non-constant optimizers across the singularity, namely, a change from unimodal to multimodal structure.

We choose ψ above for simple arguments. Though ψ is not continuous at $1/4$ in general, it will be clear that the results above also hold for a smoothed version of ψ; see the remarks below the proof of Lemma 7 in Section VI. Indeed, modified versions of the arguments in the proofs below will go through for suitable bimodal potentials ψ with a $\psi(t) = \psi(1-t)$, including potentials with a shape like the Lennard-Jones potential[21] on $[0, 1/2]$. See Figure 3.

V. DISCUSSION

An interesting connection between certain random graph models and lattice statistical mechanics models is found in the Curie-Weiss mean field Ising model. Consider the case where $d = 1$ and $\psi \equiv J$ is constant. Then the energy density becomes

$$n^{-2}J\sum_{i,j=1}^{n}\eta(i)\eta(j). \tag{5.1}$$

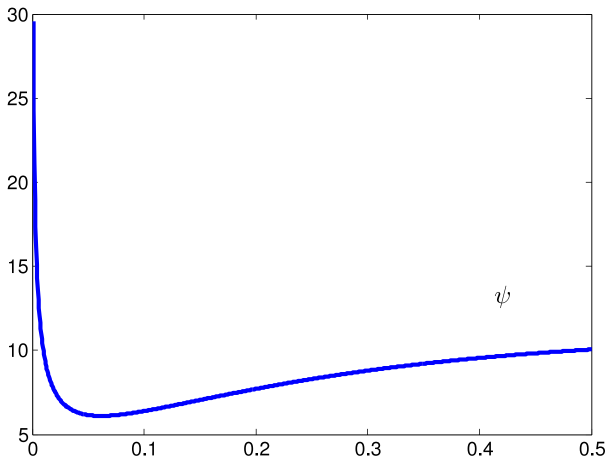

FIG. 3. An example of a smooth interaction ψ for which S is singular as in Theorem 8. Here we take periodic boundary conditions, *i.e.*, $\psi(t) = \psi(1-t)$.

This is the same as the pair interaction energy in the Curie-Weiss model in dimension $d = 1$.[20] Now consider a random graph model where a graph $X = (X_{ij})_{1\le i,j\le n}$ is represented by its adjacency matrix: $X_{ij} = 1$ if there is an edge from i to j, and $X_{ij} = 0$ otherwise. X can be directed or undirected; if it is undirected $X_{ij} = 1$ implies $X_{ji} = 1$ and vice versa. The energy density of X is defined as

$$n^{-3}\sum_{i=1}^{n}\sum_{j,k=1}^{n} X_{ij}X_{ik}. \tag{5.2}$$

Note that $X_{ij}X_{ik} = 1$ if and only if there is an edge from i to j and from i to k in X. Thus, the energy can be considered a count of the number of 2-stars embedded in X (if X is directed, the 2-stars are outward directed). In addition to the energy density, a particle density is defined as

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij}. \tag{5.3}$$

Directed and undirected versions of this model have been studied in both the grand canonical[1,29] and microcanonical[2,18,19] setting. In some cases, the outer sums (over i) in (5.2) and (5.3) can be "decoupled" from the inner sums (over j,k). The inner sums, namely, $n^{-2}\sum_{j,k=1}^{n} X_{ij}X_{ik}$ and $n^{-1}\sum_{j=1}^{n} X_{ij}$ after appropriate normalization, look like the Curie-Weiss energy and particle density in dimension $d = 1$ (when i is considered fixed), and for this reason the relevant free energies and entropies of such random graph models are closely related to the corresponding quantities in the Curie-Weiss model. See Ref. 20 for a description and analysis of the Curie-Weiss model and Refs. 1, 2, 18, 19, and 29 for details and discussion on the above mentioned random graph models. Our models differ from such random graph models in that the interaction between edges is allowed to depend on the distance between the edges.

Another way to view our models is as follows. When $d = 2$, a configuration $\eta : \{1,\ldots,n\}^2 \to \{0,1\}$ corresponds to the adjacency matrix of a directed graph: $\eta(i,j) = 1$ if there is an edge from i to j and $\eta(i,j) = 0$ otherwise. In this case, ψ corresponds to an interaction between edges. If ψ is nonconstant, it introduces a underlying geometry to the graphs. For instance, if ψ is repulsive and $\eta(i,j) = 1$, then for fixed particle density and sufficiently low energy density, other edges are not likely to appear "near" the directed edge (i,j). (For clarity, we have defined "near" in the context of the Euclidean norm. However, inspection of Lemma 11 and Lemma 12 shows that our main result, Theorem 3, continues to hold when the Euclidean norm $|\cdot|$ is replaced with any other norm.) To see how ψ might capture geometric features of graphs, consider the case of a repulsive potential with a cutoff, and assume particle density is fixed. Graphs at low energy density likely have lower connectivity, since pairs of edges at distance less than the cutoff are not likely to appear together;

on the other hand, graphs at high energy density may tend to cluster. Thus, we expect that the energy density is related to clustering and connectivity properties of the graphs. From a statistical perspective, ψ allows us to capture second as well as first order statistics of graphs, for instance, edge correlations as well as mean edge density.

When $d = 2$, the limiting occupation density is related to a certain type of graph limit called *graphon*.[22] Formally, a graphon is a symmetric measurable function $g : [0,1]^2 \to [0,1]$. Intuitively, graphons g represent an edge probability density: namely, $g(x, y)$ represents the probability for an edge between x and y, where $x, y \in [0,1]$ lie on a continuum of vertices. Interestingly, it has recently been shown that in certain random graph models where the densities of edges and certain embedded subgraphs (for instance, 2-stars, as discussed above) are held constant, the graphon g_* that optimizes entropy tends to form facets, that is, g_* is either constant or piecewise constant (up to a relabeling of vertices); see for instance Refs. 18, 27, 28, and 26. In contrast, we have shown in Corollary 5 that our optimizers f_* must be continuous (provided they satisfy the Euler-Lagrange equations (3.4)). The reason seems to be that the geometry associated with ψ enforces some regularity on the structure of the optimizers.[30]

VI. PROOFS

Before proceeding with the proofs, we introduce some terminology from large deviations theory. A family of probability measures $\mathbb{Q}_n$ on $\mathcal{T}$ is called *exponentially tight* if all compact subsets of $\mathcal{T}$ are measurable and for every $M < \infty$, there is a compact set $C \subset \mathcal{T}$ such that $\lim_{n\to\infty} n^{-1} \log \mathbb{Q}_n(\mathcal{T} \setminus C) < -M$. Two families X_n, Y_n of real-valued random variables defined on the same probability space are called *exponentially equivalent* with speed n^d if for each $\delta > 0$, the event $\{|X_n - Y_n| > \delta\}$ is measurable with $\limsup_{n\to\infty} n^{-d} \log \mathbb{P}(|X_n - Y_n| > \delta) = -\infty$. Given a topological vector space $\mathcal{T}$ over $\mathbb{R}$ and a rate function $K : \mathcal{T} \to \mathbb{R}$, an *exposing hyperplane* for $y \in \mathcal{T}$ is an element $\lambda \in \mathcal{T}^*$ such that $\langle \lambda, y\rangle - K(y) > \langle \lambda, z\rangle - K(z)$ for all $z \in \mathcal{T}$ with $z \neq y$ (here $\mathcal{T}^*$ denotes the dual space of $\mathcal{T}$ and $\langle \cdot,\cdot\rangle : \mathcal{T}^* \times \mathcal{T} \to \mathbb{R}$ the natural pairing). See Ref. 16 for details.

We restate the *contraction principle* from large deviations theory for our purposes as follows. (See Theorem 4.2.1 of Ref. 16.) Let $\mathbb{Q}_n$ be a family of probability measures on a Hausdorff topological space $\mathcal{T}$ satisfying a large deviations principle with speed n^d and rate function $K : \mathcal{T} \to \mathbb{R}$. If $F : \mathcal{T} \to \mathbb{R}$ is continuous, then the family $\tilde{\mathbb{Q}}_n$ of pushforwards of $\mathbb{Q}_n$ by F (defined by $\tilde{\mathbb{Q}}_n(A) = \mathbb{Q}_n(F^{-1}(A))$ for measurable $A \subset \mathbb{R}$) satisfies a large deviations principle with speed n^d and rate function $L : \mathbb{R} \to \mathbb{R}$,

$$L(y) := \inf_{x\in\mathcal{T} \,:\, F(x)=y} K(x).$$

We begin by proving Theorem 2. First, we need the following lemmas.

Lemma 9. For any $s \in [0,1]$,

$$\sup_{t\in\mathbb{R}} \left[st - \log\left(\frac{1}{2} + \frac{1}{2}e^t\right)\right] = H_{\text{bin}}(s). \tag{6.1}$$

Proof. When $s \notin [0,1]$ the quantity in brackets has no upper bound in t. When $s \in [0,1]$, the maximum is attained when $t = \log s - \log(1 - s)$, and plugging this back into (6.1) yields the result. □

Lemma 10. Suppose $\theta : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and let $g \in L^p(C)$. Then

$$\lim_{n\to\infty} n^{-d} \sum_{I\in\Lambda_n} \theta\left(n^d \int_{C_I} g(x)\,dx\right) = \int_C \theta(g(x))\,dx.$$

Proof. Consider the operator $A_n : L^q(C) \to L^q(C)$ defined by

$$A_n g = \sum_{I\in\Lambda_n} \left(n^d \int_{C_I} g(x)\,dx\right) 1_{C_I}.$$

Note that

$$\int_C \theta(A_n g(x))\,dx = n^{-d} \sum_{I\in\Lambda_n} \theta\left(n^d \int_{C_I} g(x)\,dx\right).$$

Since θ is Lipschitz, for a constant $c > 0$,

$$\begin{aligned}\left|\int_C \theta(A_n g(x))\,dx - \int_C \theta(g(x))\,dx\right| &\le \int_C |\theta(A_n g(x)) - \theta(g(x))|\,dx \\ &\le c\int_C |A_n g(x) - g(x)|\,dx.\end{aligned} \tag{6.2}$$

Clearly $A_n g \to g$ in norm when g is continuous. Since A_n is a bounded operator and continuous functions are dense in $L^q(C)$, we see that $A_n g \to g$ in norm for any $g \in L^q(C)$. Thus, the last expression in (6.2) vanishes as $n \to \infty$. □

Proof of Theorem 2. Recall that

$$\begin{aligned}\mathcal{Y} &:= \{f \in \mathcal{X} \,:\, f(x) \in [0,1] \text{ for a.e. } x\}, \\ \mathcal{F} &:= \{f \in \mathcal{X} \,:\, \exists \epsilon > 0 \text{ s.t. } f(x) \in [\epsilon, 1-\epsilon] \text{ for a.e. } x\}.\end{aligned} \tag{6.3}$$

We claim that $\mathcal{Y}$ is compact. Note that $\mathcal{Y}$ is closed, convex, and bounded in $L^p(C)$. Thus, by the Banach-Alaoglu theorem, $\mathcal{Y}$ is compact if $1 < p < \infty$. Since the weak topology in $L^1(C)$ is coarser than the weak topology in $L^p(C)$ for $1 < p < \infty$, the $p = 1$ case follows. We follow Baldi's theorem; see Theorem 4.5.3 of Ref. 16. Let $\mathbb{E}_n$ be expectation associated with $\mathbb{P}_n$. Write $f^{\eta_n} : C \to [0,1]$ for the function drawn from $\mathbb{P}_n$ associated with $\eta_n : \Lambda_n \to \{0,1\}$. Thus, $(\eta_n(I))_{I\in\Lambda_n}$ are i.i.d Bernoulli-1/2 random variables. For any $g \in L^q(C)$,

$$\begin{aligned}H^*(g) &:= \lim_{n\to\infty} n^{-d} \log \mathbb{E}_n\left[\exp\left(n^d \int_C f^{\eta_n}(x) g(x)\,dx\right)\right] \\ &= \lim_{n\to\infty} n^{-d} \log \mathbb{E}_n\left[\exp\left(n^d \sum_{I\in\Lambda_n} \eta_n(I) \int_{C_I} g(x)\,dx\right)\right] \\ &= \lim_{n\to\infty} n^{-d} \log \prod_{I\in\Lambda_n} \mathbb{E}_n\left[\exp\left(\eta_n(I) n^d \int_{C_I} g(x)\,dx\right)\right] \\ &= \lim_{n\to\infty} n^{-d} \sum_{I\in\Lambda_n} \log \mathbb{E}_n\left[\exp\left(\eta_n(I) n^d \int_{C_I} g(x)\,dx\right)\right] \\ &= \lim_{n\to\infty} n^{-d} \sum_{I\in\Lambda_n} \log\left[\frac{1}{2} + \frac{1}{2}\exp\left(n^d \int_{C_I} g(x)\,dx\right)\right] \\ &= \int_C \log\left(\frac{1}{2} + \frac{1}{2}e^{g(x)}\right)\,dx.\end{aligned}$$

The last equality follows from Lemma 10, since $\theta(t) := \log(\frac{1}{2} + \frac{1}{2}e^t)$ is Lipschitz. Notice $\mathcal{Y}$ is compact and the $\mathbb{P}_n$ are supported on $\mathcal{Y}$. In particular, $\mathbb{P}_n$ is exponentially tight. Thus (see Theorem 4.5.3 (a) of Ref. 16) $\mathbb{P}_n$ satisfies the large deviations upper bound in $L^p(C)$ with rate function

$$\begin{aligned}H(f) &= \int_C H_{\text{bin}}(f(x))\,dx \\ &= \sup_{g\in L^q(C)} \left[\int_C f(x)g(x)\,dx - \int_C \log\left(\frac{1}{2} + \frac{1}{2}e^{g(x)}\right)\,dx\right],\end{aligned}$$

where we used Lemma 9 for the second equality. Since the weak topology is coarser than the norm topology, $\mathbb{P}_n$ also satisfies the large deviations upper bound in $\mathcal{X}$. It is easy to check that the rate function H is nonnegative and lower semi-continuous. We now verify the remaining conditions in Baldi's theorem. Let $f \in \mathcal{F}$, and define $h_f(x) = \log f(x) - \log(1 - f(x))$ for $x \in C$. Then h_f is an exposing hyperplane for f, since

$$\int_C f(x)h_f(x)\,dx - H(f) - \left[\int_C g(x)h_f(x)\,dx - H(g)\right]$$
$$= \int_C \left(g(x)\log\frac{g(x)}{f(x)} + (1-g(x))\log\frac{1-g(x)}{1-f(x)}\right)\,dx > 0 \tag{6.4}$$

whenever $f \neq g$ on a set of positive measure. Clearly, $H^*(h_f)$ exists and $H^*(\gamma h_f)$ exists and is finite for all $\gamma > 1$. If for all open sets $U \subset \mathcal{X}$, we have

$$\inf_{f\in U\cap\mathcal{F}} H(f) = \inf_{f\in U} H(f), \tag{6.5}$$

then (see Theorem 4.5.20 (b)-(c) of Ref. 16) $\mathbb{P}_n$ satisfies the large deviations lower bound in $\mathcal{X}$. The norm topology in $L^p(C)$ is coarser than the uniform topology, since an ϵ-ball in $L^p(C)$ contains the corresponding uniform ϵ-ball when $\epsilon < 1$. Thus, to prove (6.5) it suffices to consider a set U open in the uniform topology. If $U \cap \mathcal{F} = \emptyset$ then $H(f) = \infty$ for all $f \in U$ and both sides of (6.5) equal ∞. Suppose then that $f \in U$ with $H(f) < \infty$, and define

$$f_\epsilon(x) = \begin{cases} f(x), & f(x) \in [\epsilon, 1-\epsilon] \\ 1-\epsilon, & f(x) > 1-\epsilon \\ \epsilon, & f(x) < \epsilon \end{cases}.$$

For ϵ sufficiently small, $f_\epsilon \in U \cap \mathcal{F}$ and $H(f_\epsilon) \leq H(f)$. This shows that

$$\inf_{f\in U\cap\mathcal{F}} H(f) \leq \inf_{f\in U} H(f).$$

The reverse inequality holds since $U \cap \mathcal{F} \subset U$, so we are done. □

Now we turn to the proof of Theorem 3. We will need Lemmas 11 and 12.

Lemma 11. The maps $\mathcal{Y} \to \mathbb{R}$ defined by

$$f \mapsto \int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy, \qquad f \mapsto \int_C f(x)\,dx$$

are continuous.

Proof. Let $\{f_n\}$ in $\mathcal{Y}$ converge to $f \in \mathcal{Y}$. From Assumption 1, $(x,y) \mapsto \psi(|x-y|)$ is in $L^q(C^2)$. By Jensen's inequality, it follows that $x \mapsto \int_C \psi(|x-y|)\,dy$ is in $L^q(C)$, since

$$\int_C \left(\int_C \psi(|x-y|)\,dy\right)^q dx \leq \int_{C^2} \psi(|x-y|)^q\,dx\,dy < \infty.$$

By boundedness of f, $x \mapsto \int_C f(y)\psi(|x-y|)\,dy$ is also in $L^q(C)$ and thus

$$\lim_{n\to\infty} \int_C [f_n(x) - f(x)]\left(\int_C f(y)\psi(|x-y|)\,dy\right)dx = 0. \tag{6.6}$$

Notice also that since $(x,y) \mapsto \psi(|x-y|) \in L^q(C^2)$, $y \mapsto \psi(|x-y|)$ is in $L^q(C)$. So since ψ is integrable and f_n, f are uniformly bounded, by dominated convergence

$$\lim_{n\to\infty} \int_C \left|\int_C [f_n(y) - f(y)]\psi(|x-y|)\,dy\right| dx = 0.$$

Thus, using uniform boundedness of f_n again,

$$\lim_{n\to\infty} \int_C f_n(x)\left[\int_C [f_n(y) - f(y)]\psi(|x-y|)\,dy\right] dx = 0. \tag{6.7}$$

Combining (6.6) and (6.7) yields

$$\lim_{n\to\infty} \int_{C^2} [f_n(x)f_n(y) - f(x)f(y)]\psi(|x-y|)\,dx\,dy = 0.$$

Continuity of the other map is clear, so the proof is complete. □

Next we prove exponential equivalences for the sums defining $N_n(\eta)$ and $E_n(\eta)$.

Lemma 12. For any $\epsilon > 0$,

$$\limsup_{n\to\infty} n^{-d} \log \mathbb{P}_n \left(\left| n^{-d} \sum_{I\in\Lambda_n} \eta_n(I) - \int_C f^{\eta_n}(x)\,dx \right| \geq \epsilon \right) = -\infty \tag{6.8}$$

and

$$\begin{aligned}\limsup_{n\to\infty} n^{-d} \log \mathbb{P}_n \Bigg(\Bigg| n^{-2d} & \sum_{I,J\in\Lambda_n} \eta_n(I)\eta_n(J)\phi_n(I,J) \\ & - \int_{C^2} f^{\eta_n}(x) f^{\eta_n}(y)\psi(|x-y|)\,dx\,dy \Bigg| \geq \epsilon \Bigg) = -\infty.\end{aligned} \tag{6.9}$$

Proof. By the definitions of η_n and f^{ν_n},

$$n^{-d} \sum_{I\in\Lambda_n} \eta_n(I) = \int_C f^{\eta_n}(x)\,dx,$$

which implies (6.8). Define

$$\phi_n^{I,J} = \int_{C_I\times C_J} \psi(|x-y|)\,dx\,dy$$

and observe that

$$\begin{aligned} & \left| n^{-2d} \sum_{I,J\in\Lambda_n} \eta_n(I)\eta_n(J)\phi_n(I,J) - \int_{C^2} f^{\nu_n}(x) f^{\nu_n}(y)\psi(|x-y|)\,dx\,dy \right| \\ & = \left| \sum_{I,J\in\Lambda_n} \eta_n(I)\eta_n(J)\left(n^{-2d}\phi_n(I,J) - \phi_n^{I,J}\right) \right| \\ & \leq \sum_{I,J\in\Lambda_n} \left| n^{-2d}\phi_n(I,J) - \phi_n^{I,J} \right|. \end{aligned} \tag{6.10}$$

Using Riemann integrability of $(x,y) \mapsto \psi(|x-y|)$, it is easy to see that the last expression in (6.10) is less than ϵ for sufficiently large n. This implies (6.9). □

Proof of Theorem 3. This is an immediate consequence of the contraction principle[16] along with Theorem 2, Lemma 11, and Lemma 12. □

Now we are ready to prove Theorem 4.

Proof of Theorem 4. Since $\mathcal{Y}$ is compact and

$$f \in \mathcal{Y} \mapsto \int_{C^2} f(x)f(y)\psi(|x-y|)\,dx\,dy, \qquad f \in \mathcal{Y} \mapsto \int_C f(x)\,dx$$

are continuous, $\mathcal{Y}_{\xi,\rho}$ is compact. Thus, optimizers of (3.3) exist when $(\xi,\rho) \in \Omega$. Suppose now that $f_* \in \mathcal{F}$ is an optimizer of (3.3) for some (ξ,ρ). For the remainder of the proof we will equip $\mathcal{Y}$ with the topology induced by the uniform norm. Thus, f_* is in the interior of $\mathcal{Y}$. To obtain the Euler-Lagrange equations (3.4) we follow Theorem 9.1 of Ref. 14. The multiplier rule there states that there exist $\beta,\mu \in \mathbb{R}$ and $\nu \in \{0,1\}$ such that $(\beta,\mu,\nu) \neq (0,0,0)$ and for $f = f_*$ and all $\delta f \in L^\infty(C)$,

$$0 = \beta \int_C \left(\int_C f(y)\psi(|x-y|)\,dy \right) \delta f(x)\,dx + \mu \int_C \delta f(x)\,dx - \nu \int_C H'_{\text{bin}}(f(x))\delta f(x)\,dx, \tag{6.11}$$

provided the Frechét derivatives in (6.11) are continuous for $f \in \mathcal{F}$. Continuity of the second Frechét derivative is obvious. Continuity of the first Frechét derivative follows from integrability of

$(x,y)\mapsto\psi(|x-y|)$, and continuity of the third Frechét derivative follows from uniform continuity of H'_{bin} on $[\epsilon,1-\epsilon]$ for each $\epsilon>0$. Thus,

$$\beta\int_C f_*(y)\psi(|x-y|)\,dy+\mu-\nu H'_{\text{bin}}(f_*(x))=0 \tag{6.12}$$

for a.e. x. When $\nu=1$, this is a rearrangement of (3.4). If $\nu=0$ then $\beta\neq 0$ and

$$\int_C f_*(y)\psi(|x-y|)\,dy=\gamma$$

for a.e. x, where $\gamma=-\mu/\beta$. Note that

$$\xi=\int_{C^2} f_*(x)f_*(y)\psi(|x-y|)\,dy\,dx=\gamma\int_C f_*(x)\,dx=\gamma\rho,$$

so in fact $\gamma=\xi/\rho$. □

Proof of Corollary 5. Let f satisfy (3.4). Since $\gamma(t):=\psi(|t|)$ is locally integrable,

$$(f*\gamma)(x)\equiv\int_C f(y)\psi(|x-y|)\,dy$$

is continuous. Now continuity of f follows from the Euler-Lagrange equation (3.4). □

Proof of Lemma 7. Let $0<\rho\le 1/4$ and define

$$f_1(x)=1_{[0,\rho]}(x),\quad f_2(x)\equiv\rho,\quad\text{and}\quad f_3(x)=1_{[0,\rho/2]}+1_{[1/2-\rho/2,1/2]}.$$

Then $\rho=\int_{[0,1]}f_i(x)\,dx$ for $i=1,2,3$, and

$$\begin{aligned}\xi(f_1)&=\frac{2\rho^{2-r}}{(1-r)(2-r)},\\ \xi(f_2)&=\lambda\rho^2=2\frac{4^{r-1}\rho^2}{1-r}+\frac{M\rho^2}{2},\\ \xi(f_3)&=\frac{4(\frac{\rho}{2})^{2-r}}{(1-r)(2-r)}+\frac{M\rho^2}{2}.\end{aligned} \tag{6.13}$$

Observe that when $\rho=1/4$,

$$\xi(f_2)=\frac{4^{r-5/2}}{1-r}+\frac{M\rho^2}{2}<\frac{4^{3r/2-2}}{(1-r)(2-r)}+\frac{M\rho^2}{2}=\xi(f_3), \tag{6.14}$$

where the inequality can be checked by straightforward calculus. If $\epsilon>0$ is sufficiently small, $\xi(f_2)<\xi(f_3)$ whenever $\rho\in(1/4-\epsilon,1/4)$. Moreover, when M is sufficiently large, $\xi(f_1)<\xi(f_2)$. All values of ξ between $\xi(f_1)$ and $\xi(f_3)$ are attainable by, for example, taking convex combinations of f_1 and f_3. □

To see that Lemma 7 holds for a smoothed function of ψ, let γ be a bounded function supported on $(1/4-\delta,1/4+\delta)$ such that $\gamma+\psi$ is smooth. Then for sufficiently small $\delta>0$, the arguments above still go through.

Proof of Theorem 8. Let ψ be as in Lemma 7. Take $\rho\in(1/4-\epsilon,1/4)$, and let f be an optimizer of (3.3) at $(\xi,\rho)\in\text{int}(\Omega)$. We can write $f(x)=\rho+\delta f(x)$, where

$$\int_{[0,1]}\delta f(x)\,dx=0. \tag{6.15}$$

Observe that

$$\begin{aligned}H(f)-H(\rho)&=\int_{[0,1]}H_{\text{bin}}(\rho+\delta f(x))\,dx-H_{\text{bin}}(\rho)\\ &=\int_{[0,1]}\left[H_{\text{bin}}(\rho+\delta f(x))-H'_{\text{bin}}(\rho)\delta f(x)-H_{\text{bin}}(\rho)\right]dx\\ &\ge c\int_{[0,1]}\delta f(x)^2\,dx,\end{aligned}$$

where by convexity,

$$c = \min_{t\in[-\rho,1-\rho]\setminus\{0\}} \frac{H_{\text{bin}}(\rho+t) - H'_{\text{bin}}(\rho)t - H_{\text{bin}}(\rho)}{t^2} > 0.$$

Note that $H(\rho) \le H(f)$ with equality if and only if $f \equiv \rho$ a.e. It follows that the optimizer of (3.3) at $(\lambda\rho^2, \rho)$ is the constant function with value ρ. Thus,

$$\begin{aligned}\delta S := S(\xi,\rho) - S(\lambda\rho^2,\rho) = H(\rho) - H(f)\\ \le -c\int_{[0,1]} \delta f(x)^2\,dx.\end{aligned}$$

Now note that

$$\begin{aligned}\delta\xi &:= \xi(f) - \xi(\rho)\\ &= \int_{[0,1]^2} (\rho+\delta f(x))(\rho+\delta f(y))\psi(|x-y|)\,dx\,dy - \int_{[0,1]^2} \rho^2\psi(|x-y|)\,dx\,dy\\ &= 2\rho\int_{[0,1]^2} \delta f(x)\psi(|x-y|)\,dx\,dy + \int_{[0,1]^2} \delta f(x)\delta f(y)\psi(|x-y|)\,dx\,dy\\ &= \int_{[0,1]^2} \delta f(x)\delta f(y)\psi(|x-y|)\,dx\,dy, \end{aligned} \tag{6.16}$$

with the last equality coming from (6.15) and the fact that for each $x \in [0,1]$,

$$\int_{[0,1]} \psi(|x-y|)\,dy = \lambda.$$

Since $r < 1/2$, the integral kernel Ψ defined by

$$\Psi f(y) = \int_{[0,1]} f(x)\psi(|x-y|)\,dx$$

is a Hilbert-Schmidt operator on $L^2[0,1]$. Thus,

$$\left|\int_{[0,1]^2} \delta f(x)\delta f(y)\psi(|x-y|)\,dx\,dy\right| \le \sigma\int_{[0,1]} \delta f(x)^2\,dx,$$

where σ is the spectral radius of Ψ. Putting this in (6.16) yields

$$|\delta\xi| \le \sigma\int_{[0,1]} \delta f(x)^2\,dx.$$

Combining the estimates for δS and $\delta\xi$, we get

$$\delta S \le -\frac{c}{\sigma}|\delta\xi|.$$

Thus, S is not differentiable at $(\lambda\rho^2,\rho) \in \Omega$. □

ACKNOWLEDGMENTS

The authors would like to thank R. Mark Bradley, Olivier Pinaud, Dan Pirjol, Charles Radin, and Clayton Shonkwiler for helpful comments and suggestions. D. Aristoff gratefully acknowledges support from the National Science Foundation via the Award No. NSF-DMS-1522398. L. Zhu gratefully acknowledges support from the National Science Foundation via the Award No. NSF-DMS-1613164. The authors would also like to thank an anonymous referee for a careful reading of the article and helpful suggestions.

[1] Aristoff, D. and Zhu, L., "On the phase transition curve in a directed exponential random graph model," e-print arXiv:1404.6514.
[2] Aristoff, D. and Zhu, L., "Asymptotic structure in constrained directed random graphs," Stochastic Processes Appl. **125**(11), 4154–4177 (2015).

[3] Barré, J., "Microcanonical solution of lattice models with long range interactions," Physica A **305**, 172–177 (2002).
[4] Barré, J., Bouchet, F., Dauxois, T., and Ruffo, S., "Large deviation techniques applied to systems with long-range interactions," J. Stat. Phys. **119**(314), 677–713 (2005).
[5] Barré, J., Mukamel, D., and Ruffo, S., "Inequivalence of ensembles in a system with long range interactions," Phys. Rev. Lett. **87**, 030601 (2001).
[6] Benois, O., Bodineau, T., Buttà, P., and Presutti, E., "On the validity of van der Waals theory of surface tension," Markov Processes Relat. Fields **3**, 175198 (1997).
[7] Benois, O., Bodineau, T., and Presutti, E., "Large deviations in the van der Waals limit," Stochastic Processes Appl. **75**(1), 89–104 (1998).
[8] Bouchet, F. and Barré, J., "Classification of phase transitions and ensemble inequivalence, in systems with long range interactions," J. Stat. Phys. **118**(5/6), 1073–1105 (2005).
[9] Costenuic, M., Ellis, R. S., and Touchette, H., "Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model," J. Math. Phys. **46**, 063301 (2005).
[10] Ellis, R. S., Otto, P. T., and Touchette, H., "Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model," Ann. Appl. Probab. **15**(3), 2203–2254 (2005).
[11] Campa, A., Giansanti, A., and Moroni, D., "Canonical solution of classical magnetic models with long-range couplings," J. Phys. A: Math. Gen. **36**, 6897–6921 (2003).
[12] Campa, A., Dauxois, T., and Ruffo, S., "Statistical mechanics and dynamics of solvable models with long-range interactions," Phys. Rep. **480**, 57–159 (2009).
[13] Chatterjee, S. and Diaconis, P., "Estimating and understanding exponential random graph models," Ann. Stat. **41**, 2428–2461 (2013).
[14] Clarke, F., *Functional Analysis, Calculus of Variations and Optimal Control* (Springer-Verlag, London, 2013).
[15] Dauxois, T., Ruffo, S., Arimondo, E., and Wilkens, M., *Dynamics and Thermodynamics with Long Range Interactions: An Introduction*, Lecture Notes in Physics (Springer, 2002).
[16] Dembo, A. and Zeitouni, O., *Large Deviations Techniques and Applications*, 2nd ed. (Springer, New York, 1998).
[17] Kac, M., Uhlenbeck, G. E., and Hemmer, P. C., "On the van der Waals Theory of the Vapor-Liquid Equilibrium. I. Discussion of a one dimensional Model," J. Math. Phys. **4**(2), 216–228 (1963).
[18] Kenyon, R., Radin, E., Ren, K., and Sadun, L., "Multipodal structure and phase transitions in large constrained graphs," e-print arXiv:1405.0599v2.
[19] Kenyon, R., Radin, E., Ren, K., and Sadun, L., "Bipodal structure in oversaturated random graphs," IMRN (published online).
[20] Kochmański, M., Paszkiewicz, T., and Wolsk, S., "Curie-Weiss magnet a simple model of phase transition," Eur. J. Phys. **34**(6), 1555 (2013).
[21] Lennard-Jones, J. E., "On the determination of molecular fields," Proc. R. Soc. A **106**(738), 463477 (1924).
[22] Lovász, L., *Large Networks and Graph Limits* (American Mathematical Society Colloquium Publications, 2012), Vol. 60.
[23] Mori, T., "Existence of shape-dependent thermodynamic limit in spin systems with short- and long-range interactions," J. Phys. A: Math. Theor. **48**, 145001 (2015).
[24] Mori, T., "Microcanonical analysis of exactness of the mean-field theory in long-range interacting systems," J. Stat. Phys. **147**, 1020–1040 (2012).
[25] Mori, T., "Phase transitions in systems with non-additive long-range interactions," J. Stat. Mech. **2013**, 10003.
[26] Radin, C., Ren, K., and Sadun, L., "The asymptotics of large constrained graphs," J. Phys. A: Math. Theor. **47**, 175001 (2014).
[27] Radin, C. and Sadun, L., "Phase transitions in a complex network," J. Phys. A: Math. Theor. **46**, 305002 (2013).
[28] Radin, C. and Sadun, L., "Singularities in the entropy of asymptotically large simple graphs," J. Stat. Phys. **158**, 853–865 (2015).
[29] Radin, C. and Yin, M., "Phase transitions in exponential random graphs," Ann. Appl. Probab. **23**, 2458–2471 (2013).
[30] Ruelle, D., *Statistical Mechanics: Rigorous Results* (Imperial College Press, London, 1999).
[31] Varadhan, S. R. S., *Large Deviations and Applications* (SIAM, Philadelphia, 1984).
[32] Vollmayr-Lee, B. P. and Luijten, E., "Kac-potential treatment of nonintegrable interaction," Phys. Rev. E **63**, 031108 (2001).