JOURNAL OF MATHEMATICAL PHYSICS **46**, 063301 (2005)

Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model

Marius Costeniuc[a] and Richard S. Ellis[b]
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003

Hugo Touchette[c]
School of Mathematical Sciences, Queen Mary, University of London, London, E1 4NS United Kingdom

(Received 27 October 2004; accepted 11 March 2005; published online 13 May 2005)

Using the theory of large deviations, we analyze the phase transition structure of the Curie–Weiss–Potts spin model, which is a mean-field approximation to the nearest-neighbor Potts model. It is equivalent to the Potts model on the complete graph on n vertices. The analysis is carried out both for the canonical ensemble and the microcanonical ensemble. Besides giving explicit formulas for the microcanonical entropy and for the equilibrium macrostates with respect to the two ensembles, we analyze ensemble equivalence and nonequivalence at the level of equilibrium macrostates, relating these to concavity and support properties of the microcanonical entropy. The Curie–Weiss–Potts model is the first statistical mechanical model for which such a detailed and rigorous analysis has been carried out. © *2005 American Institute of Physics.* [DOI: 10.1063/1.1904507]

I. INTRODUCTION

The nearest-neighbor Potts model, introduced in Ref. 40, takes its place next to the Ising model as one of the most versatile models in equilibrium statistical mechanics.[49] Section I C of Ref. 49 presents a mean-field approximation to the Potts model, defined in terms of a mean interaction averaged over all the sites in the model. We refer to this approximation as the Curie–Weiss–Potts model. Both the nearest-neighbor Potts model and the Curie–Weiss–Potts model are defined by sequences of probability distributions of n spin random variables that may occupy one of q different states $\theta^1, \ldots, \theta^q$, where $q \geq 3$. For $q=2$ the Potts model reduces to the Ising model while the Curie–Weiss–Potts model reduces to the much simpler mean-field approximation to the Ising model known as the Curie–Weiss model.[14]

Two ways in which the Curie–Weiss–Potts model approximates the Potts model, and in fact gives rigorous bounds on quantities in the Potts model, are discussed in Refs. 31 and 39. Probabilistic limit theorems for the Curie–Weiss–Potts model are proved in Ref. 19, including the law of large numbers and its breakdown as well as various types of central limit theorems. The model is also studied in Ref. 20, which focuses on a statistical estimation problem for two parameters defining the model.

In order to carry out the analysis of the model in Refs. 19 and 20, detailed information about the structure of the set of canonical equilibrium macrostates is required, including the fact that it exhibits a discontinuous phase transition as the inverse temperature β increases through a critical value β_c. This information plays a central role in the present paper, in which we use the theory of large deviations to study the equivalence and nonequivalence of the sets of equilibrium mac-

---

[a]Electronic mail: costeniuc@math.umass.edu
[b]Electronic mail: rsellis@math.umass.edu
[c]Electronic mail: htouchet@alum.mit.edu

rostates for the microcanonical and canonical ensembles. An important consequence of the discontinuous phase transition exhibited by the canonical ensemble in the Curie–Weiss–Potts model is the implication that the nearest-neighbor Potts model on $\mathbb{Z}^d$ also undergoes a discontinuous phase transition whenever d is sufficiently large (Ref. 4, Theorem 2.1).

In Ref. 15 the problem of the equivalence of the microcanonical and canonical ensembles was completely solved for a general class of statistical mechanical models including short-range and long-range spin models and models of turbulence. This problem is fundamental in statistical mechanics because it focuses on the appropriate probabilistic description of statistical mechanical systems. While the theory developed in Ref. 15 is complete, our understanding is greatly enhanced by the insights obtained from studying specific models. In this regard the Curie–Weiss–Potts model is an excellent choice, lying at the boundary of the set of models for which a complete analysis involving explicit formulas is available.

For the Curie–Weiss–Potts model ensemble equivalence at the thermodynamic level is studied numerically in Ref. 29, Secs. 3–5. This level of ensemble equivalence focuses on whether the microcanonical entropy is concave on its domain; equivalently, whether the microcanonical entropy and the canonical free energy, the basic thermodynamic functions in the two ensembles, can each be expressed as the Legendre–Fenchel transform of the other (Ref. 15, pp. 1036–1037). Nonconcave anomalies in the microcanonical entropy partially correspond to regions of negative specific heat and thus thermodynamic instability.

The present paper significantly extends Ref. 29, Secs. 3–5 by analyzing rigorously ensemble equivalence at the thermodynamic level and by relating it to ensemble equivalence at the level of equilibrium macrostates via the results in Ref. 15. As prescribed by the theory of large deviations, the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates and the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates are defined in (2.4) and (2.3). These macrostates are, respectively, the solutions of a constrained minimization problem involving probability vectors on $\mathbb{R}^q$ and a related, unconstrained minimization problem. The equilibrium macrostates for the two ensembles are probability vectors describing equilibrium configurations of the model in each ensemble in the thermodynamic limit $n \to \infty$. For each $i=1,2,\ldots,q$, the ith component of an equilibrium macrostate gives the asymptotic relative frequency of spins taking the spin-value θ^i.

Defined via conditioning on the energy per particle, the microcanonical ensemble expresses the conservation of physical quantities such as the energy. Among other reasons, the mathematically more tractable canonical ensemble was introduced by Gibbs[22] in the hope that in the $n \to \infty$ limit the two ensembles are equivalent; i.e., all asymptotic properties of the model obtained via the microcanonical ensemble could be realized as asymptotic properties obtained via the canonical ensemble. Although most textbooks in statistical mechanics, including Refs. 1, 22, 28, 35, 41, and 44, claim that the two ensembles always give the same predictions, in general this is not the case.[48] There are many examples of statistical mechanical models for which nonequivalence of ensembles holds over a wide range of model parameters and for which physically interesting microcanonical equilibria are often omitted by the canonical ensemble. Besides the Curie–Weiss–Potts model, these models include the mean-field Blume–Emery–Griffiths model,[2,3,18] the Hamiltonian mean-field model,[12,36] the mean-field $X-Y$ model,[11] models of turbulence,[6,16,21,33,42] models of plasmas,[34,45] gravitational systems,[23–25,37,47] and a model of the Lennard-Jones gas.[5] It is hoped that our detailed analysis of ensemble nonequivalence in the Curie–Weiss–Potts model will contribute to an understanding of this fascinating and fundamental phenomenon in a wide range of other settings.

In the present paper, after summarizing the large deviation analysis of the Curie–Weiss–Potts model in Sec. II, we give explicit formulas for the elements of $\mathcal{E}_\beta$ and the elements of $\mathcal{E}^u$ in Secs. III and IV. This analysis shows that $\mathcal{E}_\beta$ exhibits a discontinuous phase transition at a critical inverse temperature β_c and that $\mathcal{E}^u$ exhibits a continuous phase transition at a critical energy u_c. The implications of these different phase transitions concerning ensemble nonequivalence are studied graphically in Sec. V and rigorously in Sec. VI, where we exhibit a range of values of the energy u for which the microcanonical equilibrium macrostates are not realized canonically; i.e., $\mathcal{E}^u$ is disjoint from $\mathcal{E}_\beta$ for all β. As described in the main theorem in Ref. 15 and summarized here in Theorem 5.1, this range of values of the energy is precisely the set on which the microcanonical entropy is not concave. The analysis of this bridge between ensemble nonequivalence at the thermodynamic level and ensemble nonequivalence at the level of equilibrium macrostates is one of the main contributions of Ref. 15 for general models and of the present paper for the Curie–Weiss–Potts model. In a sequel to the present paper,[9] we will extend our analysis of the Curie–Weiss–Potts model to the so-called Gaussian ensemble[7,8,26,27,30,46] to show, among other results, that for each value of the energy for which the microcanonical and canonical ensembles are nonequivalent, we can find a Gaussian ensemble that is fully equivalent with the microcanonical ensemble.[10]

II. SETS OF EQUILIBRIUM MACROSTATES FOR THE TWO ENSEMBLES

Let $q \geq 3$ be a fixed integer and define $\Lambda = \{\theta^1, \theta^2, \ldots, \theta^q\}$, where the θ^i are any q distinct vectors in $\mathbb{R}^q$. In the definition of the Curie–Weiss–Potts model, the precise values of these vectors is immaterial. For each $n \in \mathbb{N}$ the model is defined by spin random variables $\omega_1, \omega_2, \ldots, \omega_n$ that take values in Λ. The canonical and microcanonical ensembles for the model are defined in terms of probability measures on the configuration spaces Λ^n, which consist of the microstates $\omega = (\omega_1, \ldots, \omega_n)$. We also introduce the n-fold product measure P_n on Λ^n with identical one-dimensional marginals

$$\bar{\rho} = \frac{1}{q}\sum_{i=1}^{q} \delta_{\theta^i}.$$

Thus for all $\omega \in \Lambda^n$, $P_n(\omega) = 1/q^n$. For $n \in \mathbb{N}$ and $\omega \in \Lambda^n$ the Hamiltonian for the q-state Curie–Weiss–Potts model is defined by

$$H_n(\omega) = -\frac{1}{2n}\sum_{j,k=1}^{n} \delta(\omega_j, \omega_k),$$

where $\delta(\omega_j, \omega_k)$ equals 1 if $\omega_j = \omega_k$ and equals 0 otherwise. The energy per particle is defined by

$$h_n(\omega) = \frac{1}{n} H_n(\omega).$$

For inverse temperature $\beta \in \mathbb{R}$ and subsets B of Λ^n the canonical ensemble is the probability measure $P_{n,\beta}$ defined by

$$P_{n,\beta}\{B\} = \frac{1}{\sum_{\omega \in \Lambda^n} \exp[-n\beta h_n(\omega)]} \cdot \sum_{\omega \in B} \exp[-n\beta h_n(\omega)].$$

For energy $u \in \mathbb{R}$ and $r > 0$ the microcanonical ensemble is the conditioned probability measure $P_n^{u,r}$ defined by

$$P_n^{u,r}\{B\} = P_n\{B | h_n \in [u-r, u+r]\}.$$

The key to our analysis of the Curie–Weiss–Potts model is to express both the canonical and the microcanonical ensembles in terms of the empirical vector

$$L_n = L_n(\omega) = (L_{n,1}(\omega), L_{n,2}(\omega), \ldots, L_{n,q}(\omega)),$$

the ith component of which is defined by

$$L_{n,i}(\omega) = \frac{1}{n}\sum_{j=1}^{n} \delta(\omega_j, \theta^i).$$

This quantity equals the relative frequency with which $\omega_j, j \in \{1, \ldots, n\}$, equals θ^i. L_n takes values in the set of probability vectors

$$\mathcal{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \ldots, \nu_q), \text{ each } \nu_i \geq 0, \sum_{i=1}^{q} \nu_i = 1 \right\}.$$

As we will see, each probability vector in $\mathcal{P}$ represents a possible equilibrium macrostate for the model.

There is a one-to-one correspondence between $\mathcal{P}$ and the set $\mathcal{P}(\Lambda)$ of probability measures on Λ, $\nu \in \mathcal{P}$ corresponding to the probability measure $\sum_{i=1}^{q} \nu_i \delta_{\theta^i}$. The element $\rho \in \mathcal{P}$ corresponding to the one-dimensional marginal $\overline{\rho}$ of the prior measures P_n is the uniform vector having equal components $1/q$.

We denote by $\langle \cdot, \cdot \rangle$ the inner product on $\mathbb{R}^q$. Since

$$\sum_{i=1}^{q} \sum_{j=1}^{n} \delta(\omega_j, \xi^i) \cdot \sum_{k=1}^{n} \delta(\omega_k, \xi^i) = \sum_{j,k=1}^{n} \delta(\omega_j, \omega_k),$$

it follows that the energy per particle can be rewritten as

$$h_n(\omega) = -\frac{1}{2n^2}\sum_{j,k=1}^{n} \delta(\omega_j, \omega_k) = -\frac{1}{2}\langle L_n(\omega), L_n(\omega)\rangle;$$

i.e.,

$$h_n(\omega) = \widetilde{H}(L_n(\omega)), \quad \text{where } \widetilde{H}(\nu) = -\tfrac{1}{2}\langle \nu, \nu \rangle \quad \text{for } \nu \in \mathcal{P}. \tag{2.1}$$

We call $\widetilde{H}$ the energy representation function.

We appeal to the theory of large deviations to define the sets of microcanonical equilibrium macrostates and canonical equilibrium macrostates. Sanov's theorem states that with respect to the product measures P_n, the empirical vectors L_n satisfy the large deviation principle (LDP) on $\mathcal{P}$ with rate function given by the relative entropy $R(\cdot|\rho)$ (Ref. 14, Theorem VIII.2.1). For $\nu \in \mathcal{P}$ this is defined by

$$R(\nu|\rho) = \sum_{i=1}^{q} \nu_i \log(q\nu_i).$$

We express this LDP by the formal notation $P_n\{L_n \in d\nu\} \approx \exp[-nR(\nu|\rho)]$. The LDPs for L_n with respect to the two ensembles $P_{n,\beta}$ and $P_n^{u,r}$ in the thermodynamic limit $n \to \infty$, $r \to 0$ can be proved from the LDP for the P_n-distributions of L_n as in Theorems 2.4 and 3.2 in Ref. 15, in which minor notational changes have to be made. We express these LDPs by the formal notation

$$P_{n,\beta}\{L_n \in d\nu\} \approx \exp[-nI_\beta(\nu)] \quad \text{and } P_n^{u,r}\{L_n \in d\nu\} \approx \exp[-nI^u(\nu)], \tag{2.2}$$

where for $\nu \in \mathcal{P}$

$$I_\beta(\nu) = R(\nu|\rho) - \frac{\beta}{2}\langle \nu, \nu \rangle - \text{const}$$

and

$$I^u(\nu) = \begin{cases} R(\nu|\rho) - \text{const} & \text{if } -\frac{1}{2}\langle \nu, \nu\rangle = u, \\ \infty & \text{otherwise}. \end{cases}$$

The constants appearing in the definitions of I_β and I^u have the properties that $\inf_{\nu\in\mathcal{P}} I_\beta(\nu)=0$ and $\inf_{\nu\in\mathcal{P}} I^u(\nu)=0$. Thus I_β and I^u map $\mathcal{P}$ into $[0,\infty)$.

As the formulas in (2.2) suggest, if $I_\beta(\nu)>0$ or $I^u(\nu)>0$, then ν has an exponentially small probability of being observed in the corresponding ensemble in the thermodynamic limit. Hence it makes sense to define the corresponding sets of equilibrium macrostates to be

$$\mathcal{E}_\beta = \{\nu \in \mathcal{P}: I_\beta(\nu) = 0\} \quad \text{and } \mathcal{E}^u = \{\nu \in \mathcal{P}: I^u(\nu) = 0\}.$$

A rigorous justification for this is given in Ref. 15, Theorem 2.4(d). Using the formulas for I_β and I^u, we see that

$$\mathcal{E}_\beta = \left\{ \nu \in \mathcal{P}: \nu \text{ minimizes } R(\nu|\rho) - \frac{\beta}{2}\langle \nu, \nu\rangle \right\} \tag{2.3}$$

and

$$\mathcal{E}^u = \left\{ \nu \in \mathcal{P}: \nu \text{ minimizes } R(\nu|\rho) \text{ subject to } -\tfrac{1}{2}\langle \nu, \nu\rangle = u \right\}. \tag{2.4}$$

Each element ν in $\mathcal{E}_\beta$ and $\mathcal{E}^u$ describes an equilibrium configuration of the model in the corresponding ensemble in the thermodynamic limit. The ith component ν_i gives the asymptotic relative frequency of spins taking the value θ^i.

The set $\mathcal{E}^u$ is defined for all u for which the constraint in the definition of I^u is satisfied for some $\nu \in \mathcal{P}$. Otherwise, $\mathcal{E}^u$ is not defined. If $\mathcal{E}^u$ is defined, then $\mathcal{E}^u$ is nonempty; if $\mathcal{E}^u$ is not defined, then we shall set $\mathcal{E}^u = \emptyset$.

The question of equivalence of ensembles at the level of equilibrium macrostates focuses on the relationships between $\mathcal{E}^u$, defined in terms of the constrained minimization problem in (2.4), and $\mathcal{E}_\beta$, defined in terms of the related, unconstrained minimization problem in (2.3). We will focus on this question in Secs. V and VI after we determine the structures of $\mathcal{E}_\beta$ and $\mathcal{E}^u$ in the next two sections.

III. FORM OF $\mathcal{E}_\beta$ AND ITS DISCONTINUOUS PHASE TRANSITION

In this section we derive the form of the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates for all $\beta \in \mathbb{R}$. This form is given in Theorem 3.1, which shows that with respect to the canonical ensemble the Curie–Weiss–Potts model undergoes a discontinuous phase transition at the critical inverse temperature

$$\beta_c = \frac{2(q-1)}{q-2} \log(q-1). \tag{3.1}$$

In order to describe the form of $\mathcal{E}_\beta$, we introduce the function ψ that maps $[0, 1]$ into $\mathcal{P}$ and is defined by

$$\psi(w) = \left(\frac{1+(q-1)w}{q}, \frac{1-w}{q}, \ldots, \frac{1-w}{q} \right); \tag{3.2}$$

the last $q-1$ components all equal $(1-w)/q$. Recalling that ρ is the uniform vector in $\mathcal{P}$ having equal components $1/q$, we see that $\rho = \psi(0)$.

Theorem 3.1: *For $\beta > 0$ let $w(\beta)$ be the largest solution of the equation*

$$w = \frac{1-e^{-\beta w}}{1+(q-1)e^{-\beta w}}. \tag{3.3}$$

The following conclusions hold.

(a) *The quantity* $w(\beta)$ *is well defined and lies in* $[0, 1]$. *It is positive, strictly increasing, and differentiable for* $\beta \in (\beta_c, \infty)$ *and satisfies* $w(\beta_c) = (q-2)/(q-1)$ *and* $\lim_{\beta\to\infty} w(\beta) = 1$.

(b) *For* $\beta \geq \beta_c$, *define* $\nu^1(\beta) = \psi(w(\beta))$ *and let* $\nu^i(\beta)$, $i=2,\ldots,q$, *denote the points in* $\mathbb{R}^q$ *obtained by interchanging the first and ith components of* $\nu^1(\beta)$. *Then the set* $\mathcal{E}_\beta$ *defined in* (2.3) *has the form*

$$\mathcal{E}_\beta = \begin{cases} \{\rho\} & \text{for } \beta < \beta_c, \\ \{\rho, \nu^1(\beta_c), \nu^2(\beta_c), \ldots, \nu^q(\beta_c)\} & \text{for } \beta = \beta_c, \\ \{\nu^1(\beta), \nu^2(\beta), \ldots, \nu^q(\beta)\} & \text{for } \beta > \beta_c. \end{cases} \tag{3.4}$$

For $\beta \geq \beta_c$, *the vectors in* $\mathcal{E}_\beta$ *are all distinct and each* $\nu^i(\beta)$ *is continuous. The vector* $\nu^1(\beta_c)$ *is given by*

$$\nu^1(\beta_c) = \psi(w(\beta_c)) = \psi\left(\frac{q-2}{q-1}\right) = \left(1 - \frac{1}{q}, \frac{1}{q(q-1)}, \ldots, \frac{1}{q(q-1)}\right); \tag{3.5}$$

the last $q-1$ *components all equal* $1/q(q-1)$.

The form of $\mathcal{E}_\beta$ for $\beta > 0$ is proved in Appendix B from a new convex-duality theorem proved in Appendix A and from the complicated calculation of the global minimum points of a related function given in Theorem 2.1 in Ref. 19. The form of $\mathcal{E}_\beta$ for $\beta \leq 0$ is also determined in Appendix B. The other assertions in Theorem 3.1 are proved in Theorem 2.1 in Ref. 19.

For $\beta > 0$ the form of $\mathcal{E}_\beta$ reflects a competition between disorder, as represented by the relative entropy $R(\nu|\rho)$, and order, as represented by the energy representation function $-\frac{1}{2}\langle \nu, \nu\rangle$. For small $\beta > 0$, $R(\nu|\rho)$ predominates. Since $R(\nu|\rho)$ attains its minimum of 0 at the unique vector ρ, we expect that for small β, $\mathcal{E}_\beta$ should contain a single vector. On the other hand, for large $\beta > 0$, $-\frac{1}{2}\langle \nu, \nu\rangle$ predominates. This function attains its minimum at $\nu^1 = (1, 0, \ldots, 0)$ and at the vectors ν^i, $i = 1, \ldots, q$, obtained by interchanging the first and ith components of ν^1. Hence we expect that for large β, $\mathcal{E}_\beta$ should contain q distinct vectors $\nu^i(\beta)$ having the property that $\nu^i(\beta) \to \nu^i$ as $\beta \to \infty$. The major surprise of the theorem is that for $\beta = \beta_c$, $\mathcal{E}_\beta$ consists of the $q+1$ distinct vectors ρ and $\nu^i(\beta_c)$ for $i = 1, 2, \ldots, q$.

The discontinuous bifurcation in the composition of $\mathcal{E}_\beta$ from 1 vector for $\beta < \beta_c$ to $q+1$ vectors for $\beta = \beta_c$ to q vectors for $\beta > \beta_c$ corresponds to a discontinuous phase transition exhibited by the canonical ensemble. In Fig. 2 in Sec. V this phase transition is shown together with the continuous phase transition exhibited by the microcanonical ensemble. The latter phase transition and the form of the set of microcanonical equilibrium macrostates are the focus of the next section.

IV. FORM OF $\mathcal{E}^u$ AND ITS CONTINUOUS PHASE TRANSITION

We now turn to the form of the set $\mathcal{E}^u$ for all $u \in [-\frac{1}{2}, -1/2q]$, which is the set of u for which $\mathcal{E}^u$ is nonempty. In the specific case $q=3$ part (c) of Theorem 4.2 gives the form of $\mathcal{E}^u$, the calculation of which is much simpler than the calculation of the form of $\mathcal{E}_\beta$. The proof is based on the method of Lagrange multipliers, which also works for general $q \geq 4$ provided the next conjecture on the form of the elements in $\mathcal{E}^u$ is valid. The validity of this conjecture has been confirmed numerically for all $q \in \{4, 5, \ldots, 10^4\}$ and all $u \in (-\frac{1}{2}, -1/2q)$ of the form $u = -\frac{1}{2} + 0.02k$, where k is a positive integer.

Conjecture 4.1: For any $q \geq 4$ *and all* $u \in (-\frac{1}{2}, -1/2q)$, *there exist* $a \neq b \in (0,1)$ *such that modulo permutations, any* $\nu \in \mathcal{E}^u$ *has the form* $(a, b, \ldots, b)$, *the last* $q-1$ *components of which all equal* b.

Parts (a) and (b) of Theorem 4.2 are proved for general $q \geq 3$. Part (c) shows that modulo permutations, for $q=3$, $\nu \in \mathcal{E}^u$ has the form $(a(u), b(u), b(u))$ and determines the precise formulas for $a(u)$ and $b(u)$. As specified in part (d), for $q \geq 4$ we can also determine the precise formula for $\nu \in \mathcal{E}^u$ provided Conjecture 4.1 is valid.

Theorem 4.2 shows that with respect to the microcanonical ensemble the Curie–Weiss–Potts model undergoes a continuous phase transition as u decreases from the critical energy value $u_c = -1/2q$. This contrast with the discontinuous phase transition exhibited by the canonical ensemble is closely related to the nonequivalence of the microcanonical and canonical ensembles for a range of u. Ensemble equivalence and nonequivalence will be explored in the next section, where we will see that it is reflected by support and concavity properties of the microcanonical entropy. An explicit formula for the microcanonical entropy is given in Theorem 4.3.

Theorem 4.2: *For $u \in \mathbb{R}$ we define $\mathcal{E}^u$ by* (2.4). *The following conclusions hold.*

(a) *For any $q \geq 3$, $\mathcal{E}^u$ is nonempty if and only if $u \in [-\frac{1}{2}, -1/2q]$. This interval coincides with the range of the energy representation function $\tilde{H}(\nu) = -\frac{1}{2}\langle \nu, \nu \rangle$ on $\mathcal{P}$.*

(b) *For any $q \geq 3$, $\mathcal{E}^{-1/2q} = \{\rho\} = \{(1/q, 1/q, \ldots, 1/q)\}$ and*

$$\mathcal{E}^{-1/2} = \{(1,0,\ldots,0),(0,1,\ldots,0),\ldots,(0,0,\ldots,1)\}.$$

(c) *Let $q=3$. For $u \in (-\frac{1}{2}, -1/2q)$, $\mathcal{E}^u$ consists of the three distinct vectors $\{\mu^1(u), \mu^2(u), \mu^3(u)\}$, where $\mu^1(u) = (a(u), b(u), b(u))$,*

$$a(u) = \frac{1 + \sqrt{2(-6u-1)}}{3} \quad \text{and } b(u) = \frac{2 - \sqrt{2(-6u-1)}}{6}. \tag{4.1}$$

The vectors $\mu^i(u)$, $i=2,3$, denote the points in $\mathbb{R}^3$ obtained by interchanging the first and the ith components of $\mu^1(u)$.

(d) *Let $q \geq 4$ and assume that Conjecture* 4.1 *is valid. Then for $u \in (-\frac{1}{2}, -1/2q)$, $\mathcal{E}^u$ consists of the q distinct vectors $\{\mu^1(u), \ldots, \mu^q(u)\}$, where $\mu^1(u) = (a(u), b(u), \ldots, b(u))$,*

$$a(u) = \frac{1 + \sqrt{(q-1)(-2qu-1)}}{q} \quad \text{and } b(u) = \frac{q-1-\sqrt{(q-1)(-2qu-1)}}{(q-1)q}.$$

The last $q-1$ components of $\mu^1(u)$ all equal $b(u)$, and the vectors $\mu^i(u)$, $i=2,\ldots,q$, denote the points in $\mathbb{R}^q$ obtained by interchanging the first and the ith components of $\mu^1(u)$.

We return to part (b) of Theorem 4.2 in order to discuss the nature of the phase transition exhibited by the microcanonical ensemble. The functions $a(u)$ and $b(u)$ given in (4.1) are both continuous for $u \in [-\frac{1}{2}, -1/2q]$ and satisfy

$$\lim_{u \to (-1/2q)^-} a(u) = \lim_{u \to (-1/2q)^-} b(u) = \frac{1}{q} = a\left(-\frac{1}{2q}\right) = b\left(-\frac{1}{2q}\right).$$

Therefore, for $i=1,\ldots,q$, $\lim_{u \to (-1/2q)^-} \mu^i(u) = \rho$. It follows that the microcanonical ensemble exhibits a continuous phase transition as u decreases from $u_c = -1/2q$, the unique equilibrium macrostate ρ for $u = u_c$ bifurcating continuously into the q distinct macrostates $\mu^i(u)$ as u decreases from its maximum value. This is rigorously true for $q=3$. Provided Conjecture 4.1 is true, it is also true for $q \geq 4$, as one easily checks using part (d) of Theorem 4.2.

Before proving Theorem 4.2, we introduce the microcanonical entropy

$$s(u) = -\inf\left\{R(\nu|\rho) : \nu \in \mathcal{P}, -\tfrac{1}{2}\langle \nu, \nu \rangle = u\right\}. \tag{4.2}$$

As we will see in the next section, this function plays a crucial role in the analysis of ensemble equivalence and nonequivalence for the Curie–Weiss–Potts model. The domain of s is the set $\text{dom}\, s = \{u \in \mathbb{R} : s(u) > -\infty\}$; for $u \notin \text{dom}\, s$, we set $s(u) = -\infty$. Since $R(\nu|\rho) < \infty$ for all $\nu \in \mathcal{P}$, dom s equals the range of $\tilde{H}(\nu) = -\frac{1}{2}\langle \nu, \nu \rangle$ on $\mathcal{P}$, which is the interval $[-\frac{1}{2}, -1/2q]$ [Theorem 4.2(a)].

Since $0 \leq R(\nu|\rho)$ for all $\nu \in \mathcal{P}$, $s(u) \in [-\infty, 0]$ for all u. The continuity of $R(\nu|\rho)$ on $\mathcal{P}$ and the compactness of the constraint set in (4.2) guarantee that for $u \in \text{dom}\, s$ the infimum in the definition of $s(u)$ is attained for some $\nu \in \mathcal{P}$. Since $R(\nu|\rho) > R(\rho|\rho) = 0$ for all $\nu \neq \rho$, it follows that s attains its maximum of 0 at the unique value $-1/2q = -\frac{1}{2}\langle \rho, \rho \rangle$.

As we have just seen, $s(-1/2q)=0$. For $u\in(-\frac{1}{2},-1/2q)$, according to parts (c) and (d) of Theorem 4.2, $\mathcal{E}^u$ consists of the unique vector $\mu^1(u)$ modulo permutations. Since for $i=2,3,\ldots,q$, $R(\mu^i(u)|\rho)=R(\mu^1(u)|\rho)$, we conclude that

$$s(u)=-R(\mu^1(u)|\rho)=-a(u)\log(qa(u))-(q-1)b(u)\log(qb(u)).$$

Finally, for $u=-\frac{1}{2}$, modulo permutations $\mathcal{E}^u$ consists of the unique vector (1, 0,…,0) [see (4.7)], and so $s(-\frac{1}{2})=-R((1,0,\ldots,0)|\rho)=-\log q$. The resulting formulas for $s(u)$ are recorded in the next theorem, where we distinguish between $q=3$ and $q\geqslant 4$.

Theorem 4.3: *We define the microcanonical entropy $s(u)$ in* (4.2). *The following conclusions hold.*

(a) $\operatorname{dom} s=[-\frac{1}{2},-1/2q]$; *for any* $u\in\operatorname{dom} s$, $u\neq -1/2q$, $s(u)<s(-1/2q)=0$; *and* $s(-\frac{1}{2})=-\log q$.

(b) *Let* $q=3$. *Then for* $u\in(-\frac{1}{2},-1/2q)=(-\frac{1}{2},-\frac{1}{6})$,

$$s(u)=-\frac{1+\sqrt{2(-6u-1)}}{3}\log(1+\sqrt{2(-6u-1)})-\frac{2-\sqrt{2(-6u-1)}}{3}\log\left(\frac{2-\sqrt{2(-6u-1)}}{2}\right). \tag{4.3}$$

(c) *Let* $q\geqslant 4$ *and assume that Conjecture* 4.1 *is valid. Then for* $u\in(-\frac{1}{2},-1/2q)$,

$$\begin{aligned} s(u)=&-\frac{1+\sqrt{(q-1)(-2qu-1)}}{q}\log(1+\sqrt{(q-1)(-2qu-1)})\\ &-\frac{q-1-\sqrt{(q-1)(-2qu-1)}}{q}\log\left(\frac{q-1-\sqrt{(q-1)(-2qu-1)}}{q-1}\right). \end{aligned} \tag{4.4}$$

We now turn to the proof of Theorem 4.2, which gives the form of $\mathcal{E}^u$. We start by proving part (a). The set $\mathcal{E}^u$ of microcanonical equilibrium macrostates consists of all $\nu\in\mathcal{P}$ that minimize the relative entropy $R(\nu|\rho)$ subject to the constraint that

$$\tilde{H}(\nu)=-\tfrac{1}{2}\langle\nu,\nu\rangle=u.$$

Let $u=-\frac{1}{2}r^2$. Since $\mathcal{P}$ consists of all non-negative vectors in $\mathbb{R}^q$ satisfying $\nu_1+\cdots+\nu_q=1$, the constraint set in the minimization problem defining $\mathcal{E}^u$ is given by

$$C(u)=C\left(-\tfrac{1}{2}r^2\right)=\left\{\nu\in\mathbb{R}^q:\nu_1\geqslant 0,\ldots,\nu_q\geqslant 0,\sum_{j=1}^q\nu_j=1,\sum_{j=1}^q\nu_j^2=r^2\right\}. \tag{4.5}$$

Geometrically, $C(-\frac{1}{2}r^2)$ is the intersection of the non-negative orthant of $\mathbb{R}^q$, the hyperplane consisting of $\nu\in\mathbb{R}^\sigma$ that satisfy $\nu_1+\cdots+\nu_q=1$, and the hypersphere in $\mathbb{R}^q$ with center 0 and radius r. Clearly, $C(u)\neq\emptyset$ if and only if u lies in the range of the energy representation function $\tilde{H}(\nu)=-\frac{1}{2}\langle\nu,\nu\rangle$ on $\mathcal{P}$. Because $0\leqslant R(\nu|\rho)<\infty$ for all $\nu\in C(u)$, the range of $\tilde{H}$ on $\mathcal{P}$ also equals the set of u for which $\mathcal{E}^u\neq\emptyset$.

The geometric description of $C(u)$ makes it straightforward to determine those values of u for which this constraint set is nonempty. The smallest value of r for which $C(-\frac{1}{2}r^2)\neq\emptyset$ is obtained when the hypersphere of radius r is tangent to the hyperplane, the point of tangency being $\rho=(1/q,1/q,\ldots,1/q)$, the closest probability vector to the origin. The hypersphere and the hyperplane are tangent when $r=1/\sqrt{q}$, which coincides with the distance from the center of the hypersphere to the hyperplane. It follows that the largest value of u for which $C(u)\neq\emptyset$, and thus $\mathcal{E}^u\neq\emptyset$, is $u=-\frac{1}{2}r^2=-1/2q$. In this case

$$C\left(-\frac{1}{2q}\right)=\{\rho\}=\left\{\left(\frac{1}{q},\frac{1}{q},\ldots,\frac{1}{q}\right)\right\}=\mathcal{E}^{-1/2q}. \tag{4.6}$$

For all sufficiently large r, $C\left(-\frac{1}{2}r^2\right)$ is empty because the hypersphere of radius r has empty intersection with the intersection of the hyperplane and the non-negative orthant of $\mathbb{R}^q$. The largest value for r for which this does not occur is found by subtracting the two equations defining the hyperplane and the hypersphere. Since each $\nu_i \in [0,1]$, it follows that

$$0 \leqslant \sum_{i=1}^{q} \nu_i(1-\nu_i) = 1 - r^2,$$

and this in turn implies that $r^2 \leqslant 1$. Thus $r=1$ is the largest value for r for which $C\left(-\frac{1}{2}r^2\right) \neq \emptyset$. We conclude that the smallest value of u for which $C(u) \neq \emptyset$, and thus $\mathcal{E}^u \neq \emptyset$, is $u=-\frac{1}{2}r^2=-\frac{1}{2}$. The set $\mathcal{E}^{-\frac{1}{2}}$ consists of the points at which the hyperplane intersects each of the positive coordinate axes; i.e.,

$$\mathcal{E}^{-1/2} = \{(1,0,\ldots,0),(0,1,\ldots,0),\ldots,(0,0,\ldots,1)\}. \tag{4.7}$$

This completes the proof of part (a) of Theorem 4.2.

For $u \in \left[-\frac{1}{2},-1/2q\right]$, we now determine the form $\mathcal{E}^u$ as specified in parts (b)–(d) of Theorem 4.2. Part (b) considers any $q \geqslant 3$ and the values $u=-1/2q$ and $u=-\frac{1}{2}$, part (c) $q=3$ and $u \in \left(-\frac{1}{2}, -1/2q\right)$, and part (d) $q \geqslant 4$ and $u \in \left(-\frac{1}{2},-1/2q\right)$. Part (b) has already been proved; for $u=-1/2q$ and $u=-\frac{1}{2}$, the sets $\mathcal{E}^u$ are given in (4.6) and (4.7).

We now consider $q \geqslant 3$ and $u \in \left(-\frac{1}{2},-1/2q\right)$. For $\nu \in \mathcal{P}$ define

$$K(\nu) = \sum_{j=1}^{q} \nu_j \quad \text{and } \tilde{H}(\nu) = -\tfrac{1}{2}\langle \nu, \nu\rangle.$$

By definition $\nu=(\nu_1,\ldots,\nu_q) \in \mathcal{E}^u$ if and only if ν minimizes $R(\nu|\rho)=\sum_{j=1}^q \nu_j \log(q\nu_j)$ subject to the constraints $K(\nu)=1$, $\tilde{H}(\nu)=u$, and $\nu_1 \geqslant 0,\ldots,\nu_q \geqslant 0$. For $u \in \left(-\frac{1}{2},-1/2q\right)$ we divide into two parts the calculation of the form of $\nu \in \mathcal{E}^u$. First we use Lagrange multipliers to solve the constrained minimization problem when $\nu_1 > 0,\ldots,\nu_1 > 0$. Then we argue that the vectors ν found via Lagrange multipliers solve the original constrained minimization problem when $\nu_1 \geqslant 0,\ldots,\nu_q \geqslant 0$.

We introduce Lagrange multipliers γ and λ. Any critical point of $R(\nu|\rho)$ subject to the constraints $K(\nu)=1$, $\tilde{H}(\nu)=u$, and $\nu_1 > 0$, $\nu_2 > 0,\ldots,\nu_q > 0$ satisfies

$$\nabla R(\nu|\rho) = \gamma \nabla K(\nu) + \lambda \nabla \tilde{H}(\nu),$$

$$K(\nu) = 1,$$

$$\tilde{H}(\nu) = u,$$

$$\nu_j > 0 \quad \text{for } j = 1,2,\ldots,q.$$

This system of equations is equivalent to

$$1 + \log(q\nu_j) = \gamma - \lambda \nu_j \quad \text{for } j = 1,2,\ldots,q, \tag{4.8}$$

$$\sum_{j=1}^{q} \nu_j = 1,$$

$$-\frac{1}{2}\sum_{j=1}^{q} \nu_j^2 = u,$$

$$\nu_j > 0 \quad \text{for } j = 1,2, \ldots ,q.$$

By the strict concavity of the logarithm, the first equation can have at most two solutions. Hence modulo permutations, there exists $n \in \{0,1,\ldots,q\}$ and distinct numbers $a,b \in (0,1)$ such that the first n components of any critical point ν all equal a and the last $q-n$ components of ν all equal b. The second and third equations in (4.8) take the form

$$na + (q-n)b = 1 \quad \text{and } na^2 + (q-n)b^2 = -2u. \tag{4.9}$$

If $n=0$, then $b=1/q$, while if $n=q$, then $a=1/q$. Both cases correspond to $\nu =(1/q,\ldots,1/q)=\rho$ and $u=-1/2q$, which does not lie in the open interval $\left(-\frac{1}{2},-1/2q\right)$ currently under consideration.

We now focus on $n \in \{1,\ldots,q-1\}$. In this case the two solutions of (4.9) are

$$a_1(n) = \frac{n - \sqrt{n(q-n)(-2qu-1)}}{nq}, \quad b_1(n) = \frac{q-n+\sqrt{n(q-n)(-2qu-1)}}{(q-n)q} \tag{4.10}$$

and

$$a_2(n) = \frac{n + \sqrt{n(q-n)(-2qu-1)}}{nq}, \quad b_2(n) = \frac{q-n-\sqrt{n(q-n)(-2qu-1)}}{(q-n)q}. \tag{4.11}$$

Since $u \in \left(-\frac{1}{2},-1/2q\right)$, these quantities are all well defined and $a_j(n) \neq b_j(n)$ for $j=1,2$. In addition,

$$a_1(q-n) = b_2(n) \quad \text{and } b_1(q-n) = a_2(n).$$

This means that the point having the first n components $a_2(n)$ and the last $q-n$ components $b_2(n)$ equals, modulo permutations, the point having the first $q-n$ components $a_1(q-n)$ and the last n components $b_1(q-n)$.

Thus, without loss of generality, we can seek solutions of the system (4.8) having the first n components $a_2(n)$ and the last $q-n$ components $b_2(n)$. While $a_2(1)$ and $b_2(1)$ are always positive for all $u \in \left(-\frac{1}{2},-1/2q\right)$, $b_2(n)$ might be negative for some $n \in \{2,\ldots,q-1\}$ and some $u \in \left(-\frac{1}{2},-1/2q\right)$. In this case the positivity constraint in the last line of (4.8) excludes such values of n and u.

We give full details when $q=3$, the case considered in part (c) of Theorem 4.2. When $q=3$, the interval $\left(-\frac{1}{2},-1/2q\right)$ equals $\left(-\frac{1}{2},-1/6\right)$ and we have $n \in \{1,2\}$. For $n=1$ and $n=2$ (4.11) takes the form

$$a_2(1) = \frac{1+\sqrt{2(-6u-1)}}{3}, \quad b_2(1) = \frac{2-\sqrt{2(-6u-1)}}{6}$$

and

$$a_2(2) = \frac{2+\sqrt{2(-6u-1)}}{6}, \quad b_2(2) = \frac{1-\sqrt{2(-6u-1)}}{3}.$$

For $u \in \left(-\frac{1}{2},-\frac{1}{4}\right)$, $b_2(2)$ is negative and hence a solution of (4.8) cannot have the form $(a_2(2),a_2(2),b_2(2))$. We conclude that when $u \in \left(-\frac{1}{2},-\frac{1}{4}\right)$, $\nu=(a_2(1),b_2(1),b_2(1))$ is, modulo permutations, the unique solution of (4.8) and thus the unique minimizer of $R(\nu|\rho)$ subject to the constraints in the last three lines of (4.8). For $u \in \left[-\frac{1}{4},-\frac{1}{6}\right)$, a straightforward calculation shows that

$$R((a_2(1),b_2(1),b_2(1))|\rho) < R((a_2(2),a_2(2),b_2(2))|\rho).$$

It follows again that $\nu=(a_2(1),b_2(1),b_2(1))$ is, modulo permutations, the unique minimizer of $R(\nu|\rho)$ subject to the constraints in the last three lines of (4.8). This completes the proof that for

$q=3$ and any $u \in \left(-\frac{1}{2}, -\frac{1}{6}\right)$, $\nu=(a_2(1), b_2(1), b_2(1))$ is, modulo permutations, the unique minimizer of $R(\mu|\rho)$ subject to the constraints $K(\nu)=1$, $\tilde{H}(\nu)=u$, and $\nu_1>0$, $\nu_2>0$, $\nu_3>0$.

We now prove for $q=3$ that the minimizers found via Lagrange multipliers when $K(\nu)=1$, $\tilde{H}(\nu)=u$, and $\nu_1>0$, $\nu_2>0$, $\nu_3>0$ also minimize $R(\nu|\rho)$ subject to the constraints $K(\nu)=1$, $\tilde{H}(\nu)=u$, and $\nu_1 \geq 0$, $\nu_2 \geq 0$, $\nu_3 \geq 0$. If $\nu=(\nu_1, \nu_2, \nu_3)$ satisfies the latter constraints and has two components equal to zero, then modulo permutations $\nu=(1,0,0)$ and $\tilde{H}(\nu)=u=-\frac{1}{2}$, which does not lie in the open interval $\left(-\frac{1}{2}, -\frac{1}{6}\right)$ currently under consideration. Thus we only have to consider the case where ν has one component equal to zero; i.e, $\nu=(0, a_0, b_0)$ with $a_0 \geq b_0$. In this case the second and third equations in (4.8) have the solution

$$a_0 = \frac{1+\sqrt{-4u-1}}{2}, \quad b_0 = \frac{1-\sqrt{-4u-1}}{2}.$$

We now claim that modulo permutations the unique minimizer of $R(\nu|\rho)$ subject to the constraints $K(\nu)=1$, $\tilde{H}(\nu)=u$, and $\nu_1 \geq 0$, $\nu_2 \geq 0$, $\nu_3 \geq 0$ has the form $(a_2(1), b_2(1), b_2(1))$ found in the preceding paragraph. The claim follows from the calculation

$$R((a_2(1),b_2(1),b_2(1))|\rho) < R((0,a_0,b_0)|\rho),$$

which is valid for all $u \in \left(-\frac{1}{2}, -\frac{1}{6}\right)$. This completes the proof of part (c) of Theorem 4.2, which gives the form of $\nu \in \mathcal{E}^u$ for $q=3$ and $u \in \left(-\frac{1}{2}, -\frac{1}{6}\right)$.

We now turn to part (d) of Theorem 4.2, which gives the form of $\mathcal{E}^u$ for $q \geq 4$ and $u \in \left(-\frac{1}{2}, -1/2q\right)$. If, as in the case $q=3$, we knew that modulo permutations, the minimizers have the form $(a, b, \ldots, b)$ as specified in Conjecture 4.1, then as in the case $q=3$ we would be able to derive explicit formulas for these minimizers. If Conjecture 4.1 is true, then it is easily verified that modulo permutations, $\mathcal{E}^u$ consists of the unique point $\nu=(a_2(1), b_2(1), \ldots, b_2(1))$, where $a_2(1)$ and $b_2(1)$ are defined in (4.11) for $u \in \left(-\frac{1}{2}, -1/2q\right)$. This gives part (d) of Theorem 4.2. The proof of the theorem is complete.

At the end of Sec. VI we will see that there exists an explicit value of $u_0 \in \left(-\frac{1}{2}, -1/2q\right)$ such that Conjecture 4.1 is valid for any $q \geq 4$ and all $u \in \left(-\frac{1}{2}, u_0\right]$. Hence for these values of u the form of $\nu \in \mathcal{E}^u$ given in part (d) of Theorem 4.2 and the formula for $s(u)$ given in part (c) of Theorem 4.3 are both rigorously true.

V. EQUIVALENCE AND NONEQUIVALENCE OF ENSEMBLES

As we saw in Sec. III, the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates undergoes a discontinuous phase transition as β increases through $\beta_c=[2(q-1)/(q-2)]\log(q-1)$, the unique macrostate ρ bifurcating discontinuously into the q distinct macrostates $\nu^{(i)}(\beta)$. By contrast, as we saw in Sec. IV, the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates undergoes a continuous phase transition as u decreases from $u_c=-1/2q$, the unique macrostate ρ bifurcating continuously into the q distinct macrostates $\mu^i(u)$. The different continuity properties of these phase transitions shows already that the canonical and microcanonical ensembles are nonequivalent. In this section we study this nonequivalence in detail and relate the equivalence and nonequivalence of these two sets of equilibrium macrostates to concavity and support properties of the microcanonical entropy s defined in (4.2). This is done with the help of Fig. 2, which is based on the form of s in Fig. 1 and on the results on ensemble equivalence and nonequivalence in Theorem 5.1. In Figs. 3 and 4 at the end of the section we give, for $q=3$, a beautiful geometric representation of $\mathcal{E}_\beta$ and $\mathcal{E}^u$ that also shows the ensemble nonequivalence for a range of u.

We start by stating in Theorem 5.1 results on ensemble equivalence and nonequivalence for the Curie–Weiss–Potts model. Theorem 5.1 summarizes Theorems 4.4, 4.6, and 4.8 in Ref. 15, which apply to a wide range of statistical mechanical models. The Curie–Weiss–Potts model is a special case. In this special case, we will show that the values of u and β in part (a)(i) of the next theorem are related by the thermodynamic formula $s'(u)=\beta$ [Theorem 6.2(b)]. For $u \in \operatorname{dom} s$ the

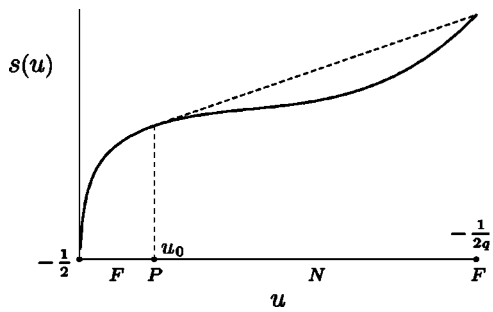

FIG. 1. Schematic graph of $s(u)$, showing the set $F=(-\frac{1}{2},u_0)\cup\{-1/2q\}$ of full ensemble equivalence, the singleton set $P=\{u_0\}$ of partial equivalence, and the set $N=(u_0,-1/2q)$ of nonequivalence. For $u\in F\cup P=(-\frac{1}{2},u_0]\cup\{-1/2q\}$, $s(u)=s^{**}(u)$; for $u\in N$, $s(u)<s^{**}(u)$ and the graph of s^{**} consists of the dotted line segment with slope β_c. The slope of s at $-\frac{1}{2}$ is ∞.

possible relationships between $\mathcal{E}^u$ and $\mathcal{E}_\beta$, given in part (a) of Theorem 5.1, are that either the ensembles are fully equivalent, partially equivalent, or nonequivalent. According to part (b) of the theorem, canonical equilibrium macrostates are always realized microcanonically—i.e., lie in $\mathcal{E}^u$ for some u—while according to part (a)(iii), microcanonical equilibrium macrostates are in general not realized canonically—i.e., do not lie in $\mathcal{E}_\beta$ for any β. It follows that the microcanonical ensemble is the richer of the two ensembles.

Theorem 5.1: *We define s by* (4.2) *and $\mathcal{E}_\beta$ and $\mathcal{E}^u$ by* (2.3) and (2.4). *The following conclusions hold.*

(a) *For fixed $u\in\operatorname{dom} s$ one of the following three possibilities occurs.*

(i) *Full equivalence: There exists $\beta\in\mathbb{R}$ such that $\mathcal{E}^u=\mathcal{E}_\beta$. This is the case if and only if s has a strictly supporting line at u with slope β; i.e.,*

$$s(v) < s(u) + \beta(v-u) \text{ for all } v \neq u.$$

(ii) *Partial equivalence: There exists $\beta\in\mathbb{R}$ such that $\mathcal{E}^u\subset\mathcal{E}_\beta$ but $\mathcal{E}^u\neq\mathcal{E}_\beta$. This is the case if and only if s has a nonstrictly supporting line at u with slope β; i.e.,*

$$s(v) \leq s(u) + \beta(v-u) \text{ for all } v\in\mathbb{R} \text{ with equality for some } v\neq u.$$

(iii) *Nonequivalence: For all $\beta\in\mathbb{R}$, $\mathcal{E}^u\cap\mathcal{E}_\beta=\varnothing$. This is the case if and only if s has no supporting line at u; i.e., for any $\beta\in\mathbb{R}$ there exists v such that $s(v)>s(u)+\beta(v-u)$.*

(b) *Canonical is always realized microcanonically: For $\nu\in\mathcal{P}$ we define $\widetilde{H}(\nu)=-\frac{1}{2}\langle\nu,\nu\rangle$. Then for any $\beta\in\mathbb{R}$,*

$$\mathcal{E}_\beta = \bigcup_{u\in\widetilde{H}(\mathcal{E}_\beta)} \mathcal{E}^u.$$

We next relate ensemble equivalence and nonequivalence with concavity and support properties of s in the Curie–Weiss–Potts model. For $q=3$ an explicit formula for s is given in part (b) of Theorem 4.3. If Conjecture 4.1 is true, then the formula for s given in part (c) of Theorem 4.3 is also valid for $q\geq 4$. Figure 1 exhibits all the concavity and support features of s. However, Fig. 1 is not the actual graph of s but a schematic graph that accentuates the shape of s together with the intervals of strict concavity and nonconcavity of s. For arbitrary $q\geq 3$, as discussed in the second paragraph after Theorem 6.2, the concavity and support features of s exhibited in Fig. 1 follow from Theorems 5.1 and 6.2.

Concavity properties of s are defined in reference to the double-Legendre–Fenchel transform s^{**}, which can be characterized as the smallest concave, upper semicontinuous function that satisfies $s^{**}(u)\geq s(u)$ for all $u\in\mathbb{R}$ (Ref. 10, Proposition A.2). For $u\in\operatorname{dom} s$ we say that s is concave at u if $s(u)=s^{**}(u)$ and that s is not concave at u if $s(u)<s^{**}(u)$. Also, we say that s is

strictly concave at $u \in \mathrm{dom}\, s$ if s has a strictly supporting line at u and that s is strictly concave on a convex subset A of $\mathrm{dom}\, s$ if s is strictly concave at each $u \in A$. If s is strictly concave at u, then a straightforward argument shows that s is concave at u, as one expects [Ref. 10, Lemma 4.1(a)].

According to Fig. 1 and Theorem 5.1, there exists $u_0 \in (-\frac{1}{2}, -1/2q)$ with the following properties:

(i) s is strictly concave on the interval $(-\frac{1}{2}, u_0)$ and at the point $-1/2q$. Hence for $u \in F = (-\frac{1}{2}, u_0) \cup \{-1/2q\}$ the ensembles are fully equivalent [Theorem. 5.1(a)(i)]. In fact, for $u \in \mathrm{int}\, F = (-\frac{1}{2}, u_0)$, $\mathcal{E}^u = \mathcal{E}_\beta$ with β given by the thermodynamic formula $\beta = s'(u)$ [Theorem 6.2(b)].

(ii) s is concave but not strictly concave at u_0 and has a nonstrictly supporting line at u_0 that also touches the graph of s over the right-hand endpoint $-1/2q$. Hence for $u = u_0$ the ensembles are partially equivalent in the sense that there exists $\beta \in \mathbb{R}$ such that $\mathcal{E}^u \subset \mathcal{E}_\beta$ but $\mathcal{E}^u \neq \mathcal{E}_\beta$ [Theorem 5.1(a)(ii)]. In fact, β equals the critical inverse temperature β_c defined in (3.1).

(iii) s is not concave on the interval $N = (u_0, -1/2q)$ and has no supporting line at any $u \in N$ [Ref. 10, Theorem A.4(c)]. Hence for $u \in N$ the ensembles are nonequivalent in the sense that for all $\beta \in \mathbb{R}$, $\mathcal{E}^u \cap \mathcal{E}_\beta = \varnothing$ [Theorem 5.1(a)(iii)].

As we have just seen, u_0 can be characterized in terms of concavity and support properties of s. The quantity u_0 can also be characterized in terms of mapping properties of $\widetilde{H}(\nu) = -\frac{1}{2}\langle \nu, \nu \rangle$. Using this characterization, we give an explicit formula for u_0 in (6.2).

We point out two additional features of Fig. 1. First, although $\mathcal{E}^u \neq \varnothing$ for u equal to the left-hand endpoint $-\frac{1}{2}$ of $\mathrm{dom}\, s$, we do not include this point in the set F of full ensemble equivalence. Indeed, s is not strictly concave at $-\frac{1}{2}$ because there is no strictly supporting line at $-\frac{1}{2}$; as one can see in (5.1), the slope of s at $-\frac{1}{2}$ is ∞. Nevertheless, by introducing the limiting set

$$\mathcal{E}_\infty = \{(1,0, \ldots ,0),(0,1, \ldots ,0), \ldots ,(0,0, \ldots ,1)\} = \lim_{\beta \to \infty} \mathcal{E}_\beta,$$

we can extend full ensemble equivalence to $u = -\frac{1}{2}$ since $\mathcal{E}^{-1/2} = \mathcal{E}_\infty$.

Second, for u in the interval N of ensemble nonequivalence, the graph of s^{**} is affine; this is depicted by the dotted line segment in Fig. 1. The slope of the affine portion of the graph of s^{**} equals the critical inverse temperature β_c defined in (3.1). This can be proved using concave-duality relationships involving s^{**} and the canonical free energy. The quantity β_c also satisfies an equal-area property, first observed by Maxwell (Ref. 28, p. 45) and explained in the context of another spin model in Ref. 18, p. 535.

The relationships stated in items (i), (ii), and (iii) above give valuable insight into equivalence and nonequivalence of ensembles in the Curie–Weiss–Potts model. These relationships are illustrated in Fig. 2. In this figure we exhibit the graph of s' and the sets $\mathcal{E}_\beta$ and $\mathcal{E}^u$ in order to compare the phase transitions in the two ensembles and to understand the implications for ensemble equivalence and nonequivalence. In order to accentuate properties of s', $\mathcal{E}_\beta$, and $\mathcal{E}^u$ that are related to ensemble equivalence and nonequivalence, we focus on $q=8$. In presenting the graph of s' and the form of $\mathcal{E}^u$, we assume that for $q=8$ Conjecture 4.1 is valid. We then appeal to part (c) of Theorem 4.3, which gives an explicit formula for s, and to part (d) of Theorem 4.2, which gives an explicit formula for the elements of $\mathcal{E}^u$. The derivative s', graphed in the top left plot in Fig. 2, is given by

$$s'(u) = \sqrt{\frac{q-1}{-2qu-1}}\left[\log(1 + \sqrt{(q-1)(-2qu-1)}) - \log\left(1 - \sqrt{\frac{-2qu-1}{q-1}}\right)\right]. \tag{5.1}$$

The canonical phase diagram, given in the top right plot in Fig. 2, summarizes the description of $\mathcal{E}_\beta$ given in Theorem 3.1 and shows the discontinuous phase transition exhibited by this ensemble at $\beta_c = [2(q-1)/(q-2)]\log(q-1) = \frac{7}{3}\log 7$. The solid line in this plot for $\beta < \beta_c$ represents the common value $\frac{1}{8}$ of each of the components of ρ, which is the unique phase for $\beta < \beta_c$. For $\beta > \beta_c$ there are eight phases given by $\nu^1(\beta)$ together with the vectors $\nu^i(\beta)$ obtained by inter-

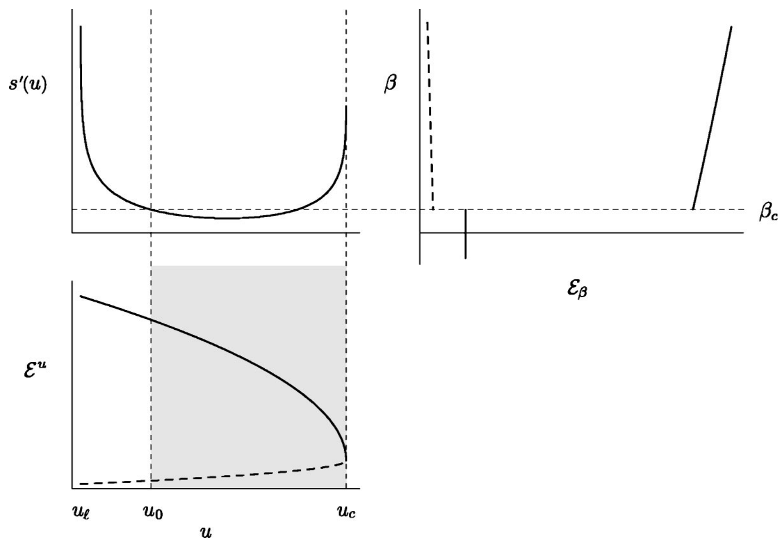

FIG. 2. For $q=8$ the top right plot shows $\mathcal{E}_\beta$, the top left plot the graph of $s'(u)$ for $u \in \operatorname{dom} s=[u_\ell, u_c]=[-\frac{1}{2}, -1/2q]$, and the bottom left plot $\mathcal{E}^u$. The discontinuous phase transition at β_c in the top right plot and the continuous phase transition at u_c in the bottom left plot imply that the ensembles are nonequivalent for all $u \in N=(u_0, u_c)$. On this interval s is not concave and s^{**} is affine with slope β_c. The shaded area in the bottom left plot corresponds to the region of nonequivalence of ensembles delineated by $u \in N$.

changing the first and ith components of $\nu^1(\beta)$. Finally, for $\beta=\beta_c$ there are nine phases consisting of ρ and the vectors $\nu^i(\beta_c)$ for $i=1,2,\ldots,8$. The solid and dashed curves in the top right plot in Fig. 2 show the first component and the last seven, equal components of $\nu^1(\beta)$ for $\beta \in [\beta_c, \infty)$. The first component is a strictly increasing function equal to $\frac{7}{8}$ for $\beta=\beta_c$ and increasing to 1 as $\beta \to \infty$ while the last seven, equal components are strictly decreasing functions equal to $\frac{1}{56}$ for $\beta=\beta_c$ and decreasing to 0 as $\beta \to \infty$.

The microcanonical phase diagram, given in the bottom left plot in Fig. 2, summarizes the description of $\mathcal{E}^u$ given in Theorem 4.2 and shows the continuous phase transition exhibited by this ensemble as u decreases from the maximum value $u_c=-1/2q=-\frac{1}{16}$. The single phase ρ for $u=-\frac{1}{16}$ is represented by the point lying over this value of u. For $u \in [-\frac{1}{2}, -\frac{1}{16})$ there are eight phases given by $\mu^1(u)$ together with the vectors $\mu^i(u)$ obtained by interchanging the first and ith components of $\mu^1(u)$. The solid and dashed curves in the bottom left plot in Fig. 2 show the first component $a(u)$ and the last seven, equal components $b(u)$ of $\mu^1(u)$ for $u \in [-\frac{1}{2}, -\frac{1}{16})$. The first component is a strictly increasing function of $-u$ equal to $\frac{1}{8}$ for $u=-\frac{1}{16}$ and increasing to 1 as $u \to (-\frac{1}{2})^+$, while the last seven, equal components are strictly decreasing functions of $-u$ equal to $\frac{1}{8}$ for $u=-\frac{1}{16}$ and decreasing to 0 as $u \to (-\frac{1}{2})^+$.

The different nature of the two phase transitions—discontinuous in the canonical ensemble versus continuous in the microcanonical ensemble—implies that the two ensembles are not fully equivalent for all values of u. By necessity, the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates must omit a set of microcanonical equilibrium macrostates. Further details concerning ensemble equivalence and nonequivalence can be seen by examining the graph of s', given in the top left plot of Fig. 2. This graph, which is the bridge between the canonical and microcanonical phase diagrams, shows that s' is strictly decreasing on the interval $\operatorname{int} F=(-\frac{1}{2}, u_0)$, which is the interior of the set F of full ensemble equivalence. The critical value β_c equals the slope of the affine portion of the graph of s^{**} over the interval $N=(u_0, -1/2q)$ of ensemble nonequivalence. This affine portion is represented in the top left plot of Fig. 2 by the horizontal dashed line at β_c.

Figure 2 exhibits the full equivalence of ensembles that holds for $u \in \operatorname{int} F=(-\frac{1}{2}, u_0)$ [Theorem 6.2(a)]. For u in this interval the solid and dashed curves representing the components of $\mu^1(u) \in \mathcal{E}^u$ can be put in one-to-one correspondence with the solid and dashed curves representing the

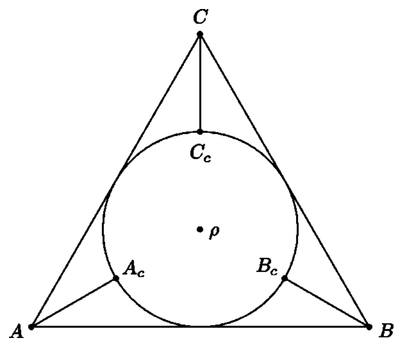

FIG. 3. Graphical representation of the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates for $q=3$ showing the maximal circle of intersection corresponding to $u=u_0$; the vector ρ; the unit-coordinate vectors A, B, and C; and the macrostates $A_c = \nu^1(\beta_c)$, $B_c=\nu^2(\beta_c)$, and $C_c=\nu^3(\beta_c)$. The line segments $\overline{A_cA}$, $\overline{B_cB}$, and $\overline{C_cC}$ represent the elements of $\mathcal{E}_\beta$ for $\beta>\beta_c$.

same two components of $\nu^1(\beta)\in\mathcal{E}_\beta$ for $\beta\in(\beta_c,\infty)$. As we remarked earlier, the values of u and β are related by the thermodynamic formula $s'(u)=\beta$ [Theorem 6.2(b)]. Full equivalence of ensembles also holds for $u=-1/2q\in F$, the right-hand endpoint of the interval on which s is finite. The solid vertical line in the top right plot for $\beta<\beta_c$, which represents the unique canonical phase ρ, is collapsed in the bottom left plot to the single energy value $u=-1/2q$, which corresponds to the unique microcanonical phase ρ. This collapse shows that the canonical notion of temperature is somewhat ill-defined at $u=-1/2q$ since there are infinitely many values of β associated with this energy value. This feature of the Curie–Weiss–Potts model is not present, for example, in the mean-field Blume–Emery–Griffiths spin model, which also exhibits nonequivalence of ensembles.[18]

By comparing the top right and bottom left plots, we see that the elements of $\mathcal{E}^u$ cease to be related to those of $\mathcal{E}_\beta$ for $u\in N=(u_0,-1/2q)$, which is the interval on which s is not concave. For any energy value u in this interval no $\nu\in\mathcal{E}_\beta$ exists that can be put in correspondence with an equivalent equilibrium empirical vector contained in $\mathcal{E}^u$. Thus, although the equilibrium macrostates corresponding to $u\in N$ are characterized by a well-defined value of the energy, it is impossible to assign an inverse temperature β to those macrostates from the viewpoint of the canonical ensemble. In other words, the canonical ensemble is blind to all energy values u contained in the interval N of nonconcavity of s. This is closely related to the presence of the discontinuous phase transition seen in the canonical ensemble.

The quantity u_0 defined in (6.2) plays a central role in the analysis of phase transitions and ensemble equivalence in the Curie–Weiss–Potts model. First, as we saw in our discussion of Fig. 1, u_0 separates the interval $\left(-\frac{1}{2},u_0\right)$ of strict concavity of s and of full ensemble equivalence from the interval $(u_0,-1/2q)$ of nonconcavity of s and of ensemble nonequivalence. Second, part (a) of Lemma 6.1 shows that u_0 equals the limiting mean energy $\widetilde{H}(\nu^1(\beta_c))$ in the canonical equilibrium macrostate $\nu^1(\beta)$ as $\beta\to(\beta_c)^+$. In Figs. 3 and 4 we present for $q=3$ a third, geometric interpretation of u_0 that is also related to ensemble nonequivalence.

Before explaining this third, geometric interpretation of u_0, we recall that according to part (a) of Theorem 4.2 specialized to $q=3$, $\mathcal{E}^u$ is nonempty, or equivalently the constraint set in (4.5) is nonempty, if and only if $u\in\left[-\frac{1}{2},-1/2q\right]=\left[-\frac{1}{2},-\frac{1}{6}\right]$. Geometrically, the energy constraint $\widetilde{H}(\nu) =-\frac{1}{2}\langle\nu,\nu\rangle=u$ corresponds to the sphere in $\mathbb{R}^3$ with center 0 and radius $\sqrt{-2u}$. This sphere intersects the set $\mathcal{P}$ of probability vectors if and only if $u\in\left[-\frac{1}{2},-\frac{1}{6}\right]$. For $u=-\frac{1}{6}$, the sphere is tangent to $\mathcal{P}$ at the unique point ρ while for $u=-\frac{1}{2}$, the hypersphere intersects $\mathcal{P}$ at the q unit-coordinate vectors. The intersection of the sphere and $\mathcal{P}$ undergoes a phase transition at u_0 in the following sense. For $u\in\left[u_0,-\frac{1}{6}\right)$ the sphere intersects $\mathcal{P}$ in a circle while for $u\in\left[-\frac{1}{2},u_0\right)$, the sphere intersects $\mathcal{P}$ in a proper subset of a circle; the complement of this subset lies outside the nonnegative octant of $\mathbb{R}^3$. For $u=u_0=-\frac{1}{4}$, the circle of intersection is maximal and is tangent to the boundary of $\mathcal{P}$.

The set $\mathcal{E}_\beta$ of canonical equilibrium macrostates for $q=3$ is represented in Fig. 3. In this figure

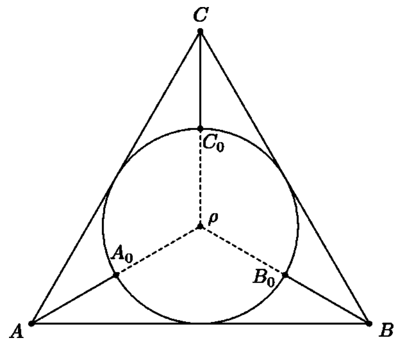

FIG. 4. Graphical representation of the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates for $q=3$ showing the maximal circle of intersection corresponding to $u=u_0$; the vector ρ; the unit-coordinate vectors A, B, and C; and the macrostates $A_0=\mu^1(u_0)$, $B_0=\mu^2(u_0)$, and $C_0=\mu^3(u_0)$. The solid-line segments $\overline{A_0A}$, $\overline{B_0B}$, and $\overline{C_0C}$ represent the elements of $\mathcal{E}^u$ that are realized canonically. The dashed-line segments $\overline{\rho A_0}$, $\overline{\rho B_0}$, and $\overline{\rho C_0}$ represent the elements of $\mathcal{E}^u$ that are not realized canonically.

the maximal circle of intersection corresponding to $u=u_0=-\frac{1}{4}$ is shown together with the vector ρ at its center; the points A, B, and C representing the respective unit-coordinate vectors $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$; and the points A_c, B_c, and C_c representing the respective equilibrium macrostates $\nu^1(\beta_c)$, $\nu^2(\beta_c)$, and $\nu^3(\beta_c)$. These three macrostates lie on the maximal circle of intersection since $\tilde{H}(\nu^1(\beta_c))=u_0$ [Lemma 6.1(b)]. For $\beta>\beta_c$ all $\nu\in\mathcal{E}_\beta$ have two equal components, and as $\beta\to\infty$ these vectors converge to the unit-coordinate vectors A, B, and C. Hence for $\beta>\beta_c$ the equilibrium macrostates $\nu^1(\beta)$, $\nu^2(\beta)$, and $\nu^3(\beta)$ are represented by the open line segments $\overline{A_cA}$, $\overline{B_cB}$, and $\overline{C_cC}$.

The set $\mathcal{E}^u$ of microcanonical equilibrium macrostates for $q=3$ is represented in Fig. 4. In this figure the maximal circle of intersection corresponding to $u=u_0=-\frac{1}{4}$ is shown together with the vector ρ at its center; the points A, B, and C representing the unit-coordinate vectors; and the points A_0, B_0, and C_0 representing the respective equilibrium macrostates $\mu^1(u_0)$, $\mu^2(u_0)$, and $\mu^3(u_0)$. For $u\in(-\frac{1}{2},-\frac{1}{6})$ all $\nu\in\mathcal{E}^u$ have two equal components, and as $u\to(-\frac{1}{2})^+$ they converge to the unit coordinate vectors A, B, and C. Hence for $u\in(-\frac{1}{2},-\frac{1}{6})$ the equilibrium macrostates $\mu^1(u)$, $\mu^2(u)$, and $\mu^3(u)$ are represented by the open line segments $\overline{\rho A}$, $\overline{\rho B}$, and $\overline{\rho C}$. As we saw in the preceding section, for each $u\in(-\frac{1}{2},-\frac{1}{6})$ the macrostates $\mu^1(u)$, $\mu^2(u)$, and $\mu^3(u)$ lie on the intersection of the sphere of radius $\sqrt{-2u}$ with $\mathcal{P}$. In particular, $A_0=\mu^1(u_0)$, $B_0=\mu^2(u_0)$, and $C_0=\mu^3(u_0)$ lie on the maximal circle of intersection.

The distinguishing feature of Fig. 4 is the three open dashed-line segments $\overline{\rho A_0}$, $\overline{\rho B_0}$, and $\overline{\rho C_0}$ representing the elements of $\mathcal{E}^u$ that are not realized canonically; namely, $\mu^1(u)$, $\mu^2(u)$, and $\mu^3(u)$ for $u\in(u_0,-\frac{1}{6})$. The three half open solid-line segments $\overline{A_0A}$, $\overline{B_0B}$, and $\overline{C_0C}$ represent the elements of $\mathcal{E}^u$ that are realized canonically; namely, $\mu^1(u)$, $\mu^2(u)$, and $\mu^3(u)$ for $u\in(-\frac{1}{2},u_0]$. For each such u the value of β for which $\mathcal{E}^u=\mathcal{E}_\beta$ is determined by the equation $\tilde{H}(\nu^1(\beta))=u$ [Theorem 6.2(a)]. Thus in Fig. 3 the corresponding elements of $\mathcal{E}_\beta$ lie on the intersection of the sphere of radius $\sqrt{-2u}$ and $\mathcal{P}$.

This completes our discussion of equivalence and nonequivalence of ensembles. In the next section we will prove a number of statements concerning ensemble equivalence and nonequivalence that have been determined graphically.

VI. PROOFS OF EQUIVALENCE AND NONEQUIVALENCE OF ENSEMBLES

Using the general results of Ref. 15, we stated in the preceding section the equivalence and nonequivalence relationships that exist between $\mathcal{E}^u$ and $\mathcal{E}_\beta$ and verified these relationships using the plots of these sets for $q=8$ given in Fig. 2. Our purpose in the present section is to prove these relationships using mapping properties of the mean energy function $u(\beta)$ defined for $\beta\neq\beta_c$ by

$$u(\beta) = \begin{cases} \tilde{H}(\rho) = -\dfrac{1}{2q} & \text{for } \beta < \beta_c, \\ \tilde{H}(\nu^1(\beta)) = -\frac{1}{2}\langle \nu^1(\beta), \nu^1(\beta)\rangle & \text{for } \beta > \beta_c. \end{cases} \tag{6.1}$$

Here $\nu^1(\beta)$ is the unique canonical equilibrium macrostate modulo permutations for $\beta > \beta_c$ [Theorem 3.1]. According to the next lemma, for $\beta > \beta_c$, $u(\beta)$ is continuous and strictly decreasing and $u(\beta) < -1/2q$, which equals the mean energy for $\beta < \beta_c$. It follows that as β increases through β_c, $u(\beta)$ is discontinuous, jumping down from $-1/2q$ to $\tilde{H}(\nu^1(\beta))$. This discontinuity in $u(\beta)$ mirrors in a natural way the discontinuity in $\mathcal{E}_\beta$ as β increases through β_c.

We use the same notation u_0 for the quantity defined in (6.2) as for the quantity u_0 appearing in Fig. 1 in Sec. V because these two quantities coincide. Indeed, with u_0 defined in (6.2), we prove in Theorem 6.2 that the largest open interval on which full equivalence of ensembles holds is int $F = \left(-\frac{1}{2}, u_0\right)$. This coincides with the interior of the interval F shown in Fig. 1. As that figure exhibits, int F is the largest open interval on which s is strictly concave; by Theorem 5.1, that open interval coincides with the largest open interval on which full equivalence of ensembles holds.

Lemma 6.1: For $\beta \in [\beta_c, \infty)$ *we define* $\nu^1(\beta)$ *as in part* (b) *of Theorem 3.1 and we define*

$$u_0 = \frac{-q^2 + 3q - 3}{2q(q-1)}. \tag{6.2}$$

The following conclusions hold.

(a) $-\frac{1}{2} < u_0 < -1/2q$ *and* $\lim_{\beta \to (\beta_c)^+} u(\beta) = \tilde{H}(\nu^1(\beta_c)) = u_0$.

(b) *The function mapping*

$$\beta \in (\beta_c, \infty) \mapsto u(\beta) = \tilde{H}(\nu^1(\beta)) = -\tfrac{1}{2}\langle \nu^1(\beta), \nu^1(\beta)\rangle$$

is a strictly decreasing, differentiable bijection onto the interval $\left(-\frac{1}{2}, u_0\right)$.

Proof: (a) The inequalities involving u_0 follow immediately from the inequality $q \geq 3$. The relationship $\tilde{H}(\nu^1(\beta_c)) = u_0$ is easily determined using the explicit form of $\nu^1(\beta_c)$ given in (3.5). That $\lim_{\beta \to (\beta_c)^+} u(\beta) = \tilde{H}(\nu^1(\beta_c))$ follows from the definition of $u(\beta)$ and the continuity of $\nu^1(\beta)$ for $\beta \geq \beta_c$.

(b) For $w \in \mathbb{R}$ define

$$f(w) = -\frac{1}{2}\left(\frac{[1 + (q-1)w]^2}{q^2} + (q-1)\frac{[1-w]^2}{q^2}\right).$$

For $\beta \in (\beta_c, \infty)$ we use the formula for $\nu^1(\beta)$ given in part (b) of Theorem 3.1 to write $u(\beta) = -f(w(\beta))$. The quantity $w(\beta)$ is positive and strictly increasing, and for all $w > 0$,

$$f'(w) = \frac{(q-1)w}{q} > 0.$$

As the composition of two strictly increasing functions, for $\beta \in (\beta_c, \infty)$, $-u(\beta)$ is strictly increasing and thus $u(\beta)$ is strictly decreasing. In addition, since $\lim_{\beta \to \infty} w(\beta) = 1$ [Theorem 3.1(a)], we have $\lim_{\beta \to \infty} u(\beta) = -\frac{1}{2}$, and by part (a) of this lemma

$$\lim_{\beta \to (\beta_c)^+} u(\beta) = \tilde{H}(\nu(\beta_c)) = u_0.$$

It follows that the function mapping $\beta \in (\beta_c, \infty) \mapsto u(\beta)$ is a strictly decreasing, differentiable bijection onto the interval $\left(-\frac{1}{2}, u_0\right)$. This completes the proof of part (b). ∎

Mapping properties of $u(\beta)$ play an important role in the next theorem, in which we prove that the sets F, P, and N defined in (6.3) correspond to full equivalence, partial equivalence, and nonequivalence of ensembles. For $u\in F$ we consider three subcases in order to indicate the value of β for which $\mathcal{E}^u=\mathcal{E}_\beta$; for $u\in \text{int}\, F=(-\frac{1}{2},u_0)$, β and u are related by $\beta=s'(u)$ and $u=u(\beta)$. Part (c) shows an interesting degeneracy in the equivalence-of-ensemble picture, the set $\mathcal{E}^u$ for $u=-1/2q$ corresponding to all $\mathcal{E}_\beta$ for $\beta<\beta_c$. This is related to the fact that for all such values of β, $\mathcal{E}_\beta=\{\rho\}$ and thus the mean energy $u(\beta)$ equals $-1/2q$.

Theorem 6.2: *We define* $s(u)$ *in* (4.2), $u(\beta)$ *in* (6.1), $\mathcal{E}_\beta$ *in* (2.3), *and* $\mathcal{E}^u$ *in* (2.4). *We also define* β_c *in* (3.1) *and* u_0 *in* (6.2). *The sets*

$$F=\left(-\tfrac{1}{2},u_0\right)\cup\left\{-\frac{1}{2q}\right\},\quad P=\{u_0\},\quad \text{and } N=\left(u_0,-\frac{1}{2q}\right) \tag{6.3}$$

have the following properties.

(a) *Full equivalence on* int F: *For* $u\in\text{int}\,F=(-\frac{1}{2},u_0)$, *there exists a unique* $\beta\in(\beta_c,\infty)$ *such that* $\mathcal{E}^u=\mathcal{E}_\beta$; β *satisfies* $u(\beta)=\tilde{H}(\nu^1(\beta))=u$.

(b) *For* $u\in\text{int}\,F=(-\frac{1}{2},u_0)$, *s is differentiable. The values u and* β *for which* $\mathcal{E}^u=\mathcal{E}_\beta$ *in part* (a) *are also related by the thermodynamic formula* $s'(u)=\beta$.

(c) *Full equivalence at* $-1/2q$: *For* $u=-(1/2q)\in F$, $\mathcal{E}^{-1/2q}=\mathcal{E}_\beta$ *for any* $\beta<\beta_c$.

(d) *Partial equivalence on P*: *For* $u\in P=\{u_0\}$, $\mathcal{E}^{u_0}\subset\mathcal{E}_{\beta_c}$ *but* $\mathcal{E}^{u_0}\neq\mathcal{E}_{\beta_c}$. *In fact,* $\mathcal{E}_{\beta_c}=\mathcal{E}^{u_0}\cup\mathcal{E}^{-1/2q}$.

(e) *Nonequivalence on N*: *For any* $u\in N=(u_0,-1/2q)$, $\mathcal{E}^u\cap\mathcal{E}_\beta=\varnothing$ *for all* $\beta\in\mathbb{R}$.

In reference to the properties of s given in part (b), one can show that the function mapping $u\in(-\frac{1}{2},u_0)\mapsto s'(u)$ is a strictly decreasing, differentiable bijection onto the interval (β_c,∞) and that this bijection is the inverse of the bijection mapping $\beta\in(\beta_c,\infty)\mapsto u(\beta)$.

Before we prove the theorem, it is instructive to compare its assertions with those in Theorem 5.1, which formulates ensemble equivalence and nonequivalence in terms of support properties of s. These support properties can be seen in the schematic plot of the the graph of s in Fig. 1. We start with part (a) of Theorem 6.2, which states that for any $u\in\text{int}\,F=(-\frac{1}{2},u_0)$ there exists a unique $\beta\in(\beta_c,\infty)$ such that $\mathcal{E}^u=\mathcal{E}_\beta$. As promised in part (a)(i) of Theorem 5.1, this β is the slope of a strictly supporting line to the graph of s at u, and so s is strictly concave on int F. The situation that holds when $u=-1/2q$ [Theorem 6.2(c)] is also consistent with part (a)(i) of Theorem 5.1. For this value of u, which is the isolated point of the set F of full equivalence, there exist infinitely many strictly supporting lines to the graph of s, the possible slopes of which are all $\beta\in(-\infty,\beta_c)$. On the other hand, when $u=u_0$, which is the only value lying in the set P of partial equivalence, we have $\mathcal{E}^{u_0}\subset\mathcal{E}_{\beta_c}$ but $\mathcal{E}^{u_0}\neq\mathcal{E}_{\beta_c}$ [Theorem 6.2(d)]. In combination with part (a)(ii) of Theorem 5.1, it follows that there exists a nonstrictly supporting line at u_0 with slope β_c and that s is concave at u_0 but not strictly concave. Finally, for $u\in N=(u_0,-1/2q)$, we have $\mathcal{E}^u\cap\mathcal{E}_\beta=\varnothing$ for all $\beta\in\mathbb{R}$ [Theorem 6.2(e)]. In accordance with part (a)(iii) of Theorem 5.1, s has no supporting line at any $u\in N$, and by Theorem A.4 in Ref. 10 s is not concave at any $u\in N$.

Proof of Theorem 6.2: (a) For $\beta>\beta_c$ part (b) of Theorem 3.1 and part (b) of Theorem 5.1 imply that

$$\mathcal{E}_\beta=\{\nu^1(\beta),\ldots,\nu^q(\beta)\}=\bigcup_{u\in\tilde{H}(\mathcal{E}_\beta)}\mathcal{E}^u.$$

The symmetry of $\tilde{H}$ with respect to permutations implies that $\tilde{H}(\mathcal{E}_\beta)=\{\tilde{H}(\nu^1(\beta))\}$. Thus for any $\beta>\beta_c$

$$\mathcal{E}_\beta=\mathcal{E}^{\tilde{H}(\nu^1(\beta))}. \tag{6.4}$$

Since for any $u\in\text{int}\,F=(-\frac{1}{2},u_0)$ there exists a unique $\beta\in(\beta_c,\infty)$ satisfying $u(\beta)=\tilde{H}(\nu^1(\beta))=u$ [Lemma 6.1(b)], it follows that $\mathcal{E}^u=\mathcal{E}_\beta$.

(b) The differentiability of s on int F is proved in part (b) of Theorem 6.3, which depends only on part (a) of the present theorem. By part (a) of the present theorem and part (a) of Theorem 5.1, s has a strictly supporting line at each $u\in\text{int}\,F$. It follows that s is strictly concave on int F and thus concave on int F [Ref. 10, Lemma 4.1(a)]; i.e., $s(u)=s^{**}(u)$ for all $u \in \text{int } F$. The differentiability of s on int F [Theorem 6.3(b)] combined with part (a) of Theorem A.3 in Ref. 10 implies that $s'(u)=\beta$.

(c) By (4.6) and part (b) of Theorem 3.1,

$$\mathcal{E}^{-1/2q} = \{\rho\} = \mathcal{E}_\beta \quad \text{for any } \beta < \beta_c. \tag{6.5}$$

(d) By part (b) of Theorem 3.1, symmetry, and part (a) of Lemma 6.1,

$$\tilde{H}(\mathcal{E}_{\beta_c}) = \{\tilde{H}(\rho), \tilde{H}(\nu^1(\beta_c))\} = \left\{-\frac{1}{2q}, u_0\right\}.$$

Hence by (6.4) and (6.5),

$$\mathcal{E}_{\beta_c} = \bigcup_{u \in \tilde{H}(\mathcal{E}_{\beta_c})} \mathcal{E}^u = \mathcal{E}^{-1/2q} \cup \mathcal{E}^{u_0} = \{\rho\} \cup \mathcal{E}^{u_0}.$$

However, $\rho \notin \mathcal{E}^{u_0}$ since ρ does not satisfy the constraint $\tilde{H}(\rho)=u_0$. It follows that $\mathcal{E}^{u_0} \subset \mathcal{E}_{\beta_c}$ but that $\mathcal{E}^{u_0} \neq \mathcal{E}_{\beta_c}$.

(e) If $u \in N$, then $u \notin (-\frac{1}{2}, u_0)$, and so by part (b) of Lemma 6.1 $u \neq \tilde{H}(\nu^1(\beta))$ for any $\beta \in (\beta_c, \infty)$. Since by (6.4) $\mathcal{E}_\beta = \mathcal{E}^{\tilde{H}(\nu^1(\beta))}$ for all $\beta > \beta_c$, it follows that for all $\beta > \beta_c$,

$$\mathcal{E}^u \cap \mathcal{E}^{\tilde{H}(\nu^1(\beta))} = \emptyset$$

and thus that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$. For any $\beta < \beta_c$ (6.5) states that $\mathcal{E}_\beta = \mathcal{E}^{-1/2q} = \{\rho\}$. Since $u \in N$, we have $u \neq -1/2q$ and thus $\mathcal{E}^{-1/2q} \cap \mathcal{E}^u = \emptyset$. It follows that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for any $\beta < \beta_c$. Finally, for $\beta = \beta_c$ part (b) of Theorem 3.1 states that $\mathcal{E}_{\beta_c} = \{\rho, \nu^1(\beta_c), \ldots, \nu^q(\beta_c)\}$. However, since $\tilde{H}(\rho) = -(1/2q) \notin N$ and $\tilde{H}(\nu^i(\beta_c)) = u_0 \notin N$, none of the vectors in $\mathcal{E}_{\beta_c}$ satisfies the constraint $\tilde{H}(\nu)=u$. Thus $\mathcal{E}^u \cap \mathcal{E}_{\beta_c} = \emptyset$. We have proved $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \in \mathbb{R}$. The proof of the theorem is complete. ■

We end this section by showing that for arbitrary $q \geq 4$ and u in the equivalence sets $F \cup P = (-\frac{1}{2}, u_0] \cup \{-1/2q\}$ the formulas for $\mathcal{E}^u$ and $s(u)$ given in part (d) of Theorem 4.2 and part (c) of Theorem 4.3 are rigorously true. Our strategy is to use the equivalence of the microcanonical and canonical ensembles for $u \in F \cup P$ and the fact that the form of $\mathcal{E}_\beta$ is known exactly for all β. Thus, we translate the form of $\nu \in \mathcal{E}_\beta$, as given in part (b) of Theorem 3.1, into the form of $\nu \in \mathcal{E}^u$ for $u \in F \cup P$. For $\beta \in [\beta_c, \infty)$, the last $q-1$ components of $\nu^1(\beta) \in \mathcal{E}_\beta$ are given by

$$\nu_j^1(\beta) = \frac{1 - w(\beta)}{q}, \tag{6.6}$$

and these components are not equal to the first component. Since for each $u \in F \cup P$ there exists $\beta \in [\beta_c, \infty]$ such that either $\mathcal{E}^u = \mathcal{E}_\beta$ or $\mathcal{E}^u \subset \mathcal{E}_\beta$, it follows that modulo permutations all $\nu \in \mathcal{E}^u$ have their last $q-1$ components equal to each other. That is, modulo permutations there exist numbers a and b in $[0, 1]$ such that $\nu = (a, b, \ldots, b)$. The possible values of a and b are easily determined by considering the constraints satisfied by $\nu \in \mathcal{E}^u$. These constraints are

$$a + (q-1)b = 1 \quad \text{and } a^2 + (q-1)b^2 = -2u.$$

The two solutions of these equations are

$$a_1 = \frac{1 - \sqrt{(q-1)(-2qu-1)}}{q}, \quad b_1 = \frac{q - 1 + \sqrt{(q-1)(-2qu-1)}}{(q-1)q}$$

and

$$a_2 = \frac{1+\sqrt{(q-1)(-2qu-1)}}{q}, \quad b_2 = \frac{q-1-\sqrt{(q-1)(-2qu-1)}}{(q-1)q}.$$

Of the two values b_1 and b_2, only b_2 has the form given in (6.6) with

$$w(\beta) = \frac{\sqrt{(q-1)(-2qu-1)}}{q-1} \in [0,1].$$

We conclude that modulo permutations each $\nu \in \mathcal{E}^u$ has the form $(a_2, b_2, \ldots, b_2)$, in which the last $q-1$ components all equal b_2. This coincides with the formula for $\mu^1(u)$ given in part (d) of Theorem 4.2, which in turn gives the explicit formula for $s(u)$ in part (c) of Theorem 4.3. This information is summarized in part (a) of the next theorem. The differentiability of s on $\operatorname{int} F$, which is stated in part (b), is an immediate consequence of the explicit formula for $s(u)$.

Theorem 6.3: *We define u_0 in* (6.2). *The following conclusions hold.*

(a) *For arbitrary $q \geqslant 4$ and u in the equivalence sets $F \cup P = (-\frac{1}{2}, u_0] \cup \{-1/2q\}$ the formulas for $\mathcal{E}^u$ and $s(u)$ given in part* (d) *of Theorem* 4.2 *and part* (c) *of Theorem* 4.3 *are rigorously true.*

(b) *For arbitrary $q \geqslant 4$, s is differentiable on the interval* $\operatorname{int} F = (-\frac{1}{2}, u_0)$ *and $s'(u)$ is given by* (5.1).

ACKNOWLEDGMENTS

The research by two of the authors (M.C. and R.S.E.) was supported by a grant from the National Science Foundation (NSF-DMS-0202309). The research of one of the authors (H.T.) was supported by the Natural Sciences and Engineering Research Council of Canada and the Royal Society of London (Canada–UK Millennium Fellowship).

APPENDIX A: TWO RELATED MAXIMIZATION PROBLEMS

Theorem A.1 is a new result on the maximum points of certain functions related by convex duality. It is formulated for a finite, differentiable, convex function F on $\mathbb{R}^\sigma$ and its Legendre–Fenchel transform,

$$F^*(z) = \sup_{x \in \mathbb{R}^\sigma} \{\langle x,z\rangle - F(x)\}.$$

The domain of F^* is the set $\operatorname{dom} F^* = \{z \in \mathbb{R}^\sigma : F^*(z) < \infty\}$. With only minor changes in notation the theorem is also valid for a finite, Gateaux-differentiable, convex function on a Hilbert space.

Theorem A.1 will be applied in Appendix B to prove that for $\beta > 0$, $\mathcal{E}_\beta$ has the form given in part (b) of Theorem 3.1. Another application of Theorem A.1 is given in Proposition 3.4 in Ref. 17. It is used there to determine the form of the set of canonical equilibrium macrostates for another important spin system known as the mean-field Blume–Emery–Griffiths model.

Theorem A.1: *Let σ be a positive integer and F a finite, differentiable, convex function mapping $\mathbb{R}^\sigma$ into $\mathbb{R}$. Assume that $\sup_{z \in \mathbb{R}^\sigma}\{F(z) - \frac{1}{2}\|z\|^2\} < \infty$ and that $F(z) - \frac{1}{2}\|z\|^2$ attains its supremum. The following conclusions hold:*

(a) $\sup_{z \in \mathbb{R}^\sigma}\{F(z) - \frac{1}{2}\|z\|^2\} = \sup_{z \in \operatorname{dom} F^*}\{\frac{1}{2}\|z\|^2 - F^*(z)\}$.
(b) $\frac{1}{2}\|z\|^2 - F^*(z)$ *attains its supremum on* $\operatorname{dom} F^*$,
(c) *the global maximum points of $F(z) - \frac{1}{2}\|z\|^2$ coincide with the global maximum points of* $\frac{1}{2}\|z\|^2 - F^*(z)$.

Proof: We define the subdifferential of F^* at $z_0 \in \mathbb{R}^\sigma$ by

$$\partial F^*(z_0) = \{y \in \mathbb{R}^\sigma : F^*(z) \geqslant F^*(z_0) + \langle y, z - z_0\rangle \text{ for all } z \in \mathbb{R}^\sigma\}.$$

We also define the domain of ∂F^* to be the set of $z_0 \in \mathbb{R}^\sigma$ for which $\partial F^*(z_0) \neq \emptyset$. The proof of the theorem uses three properties of Legendre–Fenchel transforms (see Ref. 43 for background).

(1) F^* is a convex, lower semicontinuous function mapping $\mathbb{R}^\sigma$ into $\mathbb{R}\cup\{\infty\}$, and for all $z \in \mathbb{R}^\sigma$, $F^{**}(z)=(F^*)^*(z)$ equals $F(z)$ [Ref. 14, Theorem VI.5.3(a),(e)].
(2) If for some $z_0 \in \mathbb{R}^\sigma$ and $z \in \mathbb{R}^\sigma$ we have $z=\nabla F(z_0)$, then $F(z_0)+F^*(z)=\langle z_0,z\rangle$ [Ref. 14, Theorems VI.3.5(d) and VI.5.3(c)], and so $z \in \operatorname{dom} F^*$. In particular, if $z=z_0$, then $z_0 \in \operatorname{dom} F^*$ and $F(z_0)+F^*(z_0)=\|z_0\|^2$.
(3) For $z_0 \in \operatorname{dom} F^*$ and $y \in \partial F^*(z_0)$ we have $F(y)+F^*(z_0)=\langle y,z_0\rangle$ [Ref. 14, Theorem VI.5.3(c),(d)]. In particular, if $y=z_0$, then $F(z_0)+F^*(z_0)=\|z_0\|^2$.

We first prove part (a), which is a special case of Theorem C.1 in Ref. 13. Let $M=\sup_{z\in\mathbb{R}^\sigma}\{F(z)-\|z\|^2/2\}$. Since for any $z \in \operatorname{dom} F^*$ and x in $\mathbb{R}^\sigma$,

$$F^*(z) + M \geqslant \langle x,z\rangle - F(x) + M \geqslant \langle x,z\rangle - \|x\|^2/2,$$

we have

$$F^*(z) + M \geqslant \sup_{x\in\mathbb{R}^\sigma}\{\langle x,z\rangle - \|x\|^2/2\} = \|z\|^2/2.$$

It follows that $M\geqslant\|z\|^2/2-F^*(z)$ and thus that $M\geqslant\sup_{z\in\operatorname{dom} F^*}\{\|z\|^2/2-F^*(z)\}$. To prove the reverse inequality, let $N=\sup_{z\in\operatorname{dom} F^*}\{\|z\|^2/2-F^*(z)\}$. Then for any $z \in \mathbb{R}^\sigma$ and $x \in \operatorname{dom} F^*$

$$\|z\|^2/2 + N \geqslant \langle x,z\rangle - \|x\|^2/2 + N \geqslant \langle x,z\rangle - F^*(x).$$

Since $F^*(x)=\infty$ for $x \notin \operatorname{dom} F^*$, it follows from property 1 that

$$\|z\|^2/2 + N \geqslant \sup_{x\in\operatorname{dom} F^*}\{\langle x,z\rangle - F^*(x)\} = F(z)$$

and thus that $N\geqslant\sup_{z\in\mathbb{R}^\sigma}\{F(z)-\|z\|^2/2\}$.

In order to prove parts (b) and (c) of Theorem A.1, let z_0 be any point in $\mathbb{R}^\sigma$ at which $F(z)-\frac{1}{2}\|z\|^2$ attains its supremum. Then $z_0=\nabla F(z_0)$, and so by the last line of property (2), $z_0 \in \operatorname{dom} F^*$ and $F(z_0)+F^*(z_0)=\|z_0\|^2$. Part (a) now implies that

$$\sup_{z\in\mathbb{R}^\sigma}\left\{F(z) - \tfrac{1}{2}\|z\|^2\right\} = F(z_0) - \tfrac{1}{2}\|z_0\|^2 = \tfrac{1}{2}\|z_0\|^2 - F^*(z_0) = \sup_{z\in\operatorname{dom} F^*}\left\{\tfrac{1}{2}\|z\|^2 - F^*(z)\right\}.$$

We conclude that $\frac{1}{2}\|z\|^2-F^*(z)$ attains its supremum on $\operatorname{dom} F^*$ at z_0. Not only have we proved part (b), but also we have proved half of part (c); namely, any global maximizer of $F(z)-\frac{1}{2}\|z\|^2$ is a global maximizer of $\frac{1}{2}\|z\|^2-F^*(z)$.

Now let z_0 be any point at which $\frac{1}{2}\|z\|^2-F^*(z)$ attains its supremum. Then for any $z \in \mathbb{R}^\sigma$

$$\tfrac{1}{2}\langle z_0,z_0\rangle - F^*(z_0) \geqslant \tfrac{1}{2}\langle z,z\rangle - F^*(z).$$

It follows that for any $z \in \mathbb{R}^\sigma$,

$$F^*(z) \geqslant F^*(z_0) + \tfrac{1}{2}(\langle z,z\rangle - \langle z_0,z_0\rangle) \geqslant F^*(z_0) + \langle z_0,z - z_0\rangle$$

and thus that $z_0 \in \partial F^*(z_0)$. By the last line of property (3) this implies that $F(z_0)+F^*(z_0)=\|z_0\|^2$. In conjunction with part (a) this in turn implies that

$$\sup_{z\in\operatorname{dom} F^*}\left\{\tfrac{1}{2}\|z\|^2 - F^*(z)\right\} = \tfrac{1}{2}\|z_0\|^2 - F^*(z_0) = F(z_0) - \tfrac{1}{2}\|z_0\|^2 = \sup_{z\in\mathbb{R}^\sigma}\left\{F(z) - \tfrac{1}{2}\|z\|^2\right\}.$$

We conclude that $F(z)-\frac{1}{2}\|z\|^2$ attains its supremum at z_0. This completes the proof of the theorem. ■

APPENDIX B: FORM OF $\mathcal{E}_\beta$

We first derive the form of $\mathcal{E}_\beta$ for $\beta>0$ as given in part (b) of Theorem 3.1. We then prove that $\mathcal{E}_\beta=\{\rho\}$ for all $\beta\leqslant 0$.

$\mathcal{E}_\beta$ is defined as the set of $\nu \in \mathcal{P}$ that minimize $R(\nu|\rho) - (\beta/2)\langle \nu, \nu \rangle$. Since $\beta > 0$, this is equivalent to

$$\mathcal{E}_\beta = \left\{ \nu \in \mathcal{P} : \nu \text{ maximizes } \frac{1}{2}\langle \nu, \nu \rangle - \frac{1}{\beta} R(\nu|\rho) \right\}. \tag{B1}$$

This maximization problem has the form of the right-hand side of part (a) of Theorem A.1; viz.,

$$\sup_{\nu \in \mathcal{P}} \left\{ \frac{1}{2}\langle \nu, \nu \rangle - \frac{1}{\beta} R(\nu|\rho) \right\} = \sup_{\nu \in \operatorname{dom} F^*} \left\{ \frac{1}{2}\|\nu\|^2 - F^*(\nu) \right\}$$

with $F^*(\nu) = (1/\beta) R(\nu|\rho)$.

In order to determine the function F having this Legendre–Fenchel transform, for $z \in \mathbb{R}^q$ we define the finite, differentiable, convex function

$$\Gamma(z) = \log\left(\sum_{i=1}^{q} e^{z_i} \frac{1}{q} \right) \tag{B2}$$

and set $\Gamma_\beta(z) = (1/\beta)\Gamma(\beta z)$. Since for $\nu \in \mathbb{R}^q$ (Ref. 14, Theorem VIII.2.2),

$$\Gamma^*(\nu) = \begin{cases} R(\nu|\rho) & \text{for } \nu \in \mathcal{P}, \\ \infty & \text{otherwise}, \end{cases}$$

it follows that for $\nu \in \mathbb{R}^q$,

$$(\Gamma_\beta)^*(\nu) = \sup_{z \in \mathbb{R}^q} \left\{ \langle z, \nu \rangle - \frac{1}{\beta}\Gamma(\beta z) \right\} = \frac{1}{\beta}\Gamma^*(\nu) = \begin{cases} \dfrac{1}{\beta} R(\nu|\rho) & \text{for } \nu \in \mathcal{P}, \\ \infty & \text{otherwise}. \end{cases}$$

Thus $F(z) = (1/\beta)\Gamma(\beta z)$. By part (a) of Theorem A.1,

$$\sup_{z \in \mathbb{R}^q} \left\{ \frac{1}{\beta}\Gamma(\beta z) - \frac{1}{2}\|z\|^2 \right\} = \sup_{\nu \in \mathcal{P}} \left\{ \frac{1}{2}\langle \nu, \nu \rangle - \frac{1}{\beta} R(\nu|\rho) \right\},$$

and by part (b) of the theorem the global maximum points of $\Gamma(\beta z) - \frac{1}{2}\|z\|^2$ and $\frac{1}{2}\langle \nu, \nu \rangle - (1/\beta) R(\nu|\rho)$ coincide.

Equation (B1) now implies that

$$\mathcal{E}_\beta = \left\{ z \in \mathbb{R}^q : z \text{ maximizes } \frac{1}{\beta}\Gamma(\beta z) - \frac{1}{2}\|z\|^2 \right\} = \left\{ z \in \mathbb{R}^q : z \text{ minimizes } \frac{\beta}{2}\|z\|^2 - \Gamma(\beta z) \right\}.$$

We summarize this discussion in the following corollary. Part (b) of the corollary is proved in part (b) of Theorem 2.1 in Ref. 19.

Corollary B.1: We define the finite, convex, continuous function Γ *in* (B2). *The following conclusions hold.*

(a) $\mathcal{E}_\beta$ *coincides with the set of global minimum points of*

$$G_\beta(z) = \frac{\beta}{2}\|z\|^2 - \log \sum_{i=1}^{q} e^{\beta z_i} = \frac{\beta}{2}\|z\|^2 - \Gamma(\beta z) - \log q.$$

(b) *For* $0 < \beta < \beta_c$, $\beta = \beta_c$, *and* $\beta > \beta_c$ *the set of global minimum points of* G_β *has the form given by the right-hand side of* (3.4) [*Theorem* 3.1(b)].

Corollary B.1 completes the proof of Theorem 3.1. Kiessling's proof of this corollary based on Lagrange multipliers is given in Appendix B of Ref. 20. Continuous analogues of the corollary are mentioned in Refs. 32, 33, and 38, but are not proved there.

We now show that for all $\beta \leqslant 0$, $\mathcal{E}_\beta = \{\rho\}$. This is obvious for $\beta = 0$ since $\nu = \rho$ is the unique vector in $\mathcal{P}$ that minimizes $R(\nu|\rho)$. Our goal is to prove that for $\beta < 0$, $\nu = \rho$ is also the unique vector in $\mathcal{P}$ that minimizes $R(\nu|\rho) - (\beta/2)\langle \nu, \nu\rangle$. Let $\overline{\nu}$ be a point in $\mathcal{P}$ at which $R(\nu|\rho) - (\beta/2) \times \langle \nu, \nu \rangle$ attains its infimum. For any $i = 1, 2, \ldots, q$,

$$\frac{\partial\left(R(\nu|\rho) - \dfrac{\beta}{2}\langle \nu, \nu\rangle\right)}{\partial \nu_i} = \log \nu_i + 1 - \beta \nu_i,$$

which is negative for all sufficiently small $\nu_i > 0$. It follows that $\overline{\nu}$ does not lie on the relative boundary of $\mathcal{P}$; i.e., $\overline{\nu_j} > 0$ for all $i = 1, 2, \ldots, q$. We complete the proof by showing that for any $1 \leqslant j < k \leqslant q$, $\overline{\nu_j} = \overline{\nu_k}$. Since ρ is the only point in $\mathcal{P}$ satisfying these equalities, we will be done.

Given $a \in (0,1)$, we consider the reduced two-variable problem of minimizing $R(\nu|\rho) - (\beta/2)\langle \nu, \nu\rangle$ over $\nu_j > 0$ and $\nu_k > 0$ under the constraint $\nu_j + \nu_k = a$; all the other components ν_i are fixed and equal $\overline{\nu_i}$. Setting $\nu_k = a - \nu_j$, we define

$$F(\nu_j) = R(\nu|\rho) - \frac{\beta}{2}\langle \nu, \nu\rangle.$$

Differentiating with respect to ν_j shows that any global minimizer ν_j must satisfy

$$F'(\nu_j) = \log \nu_j - \log(a - \nu_j) - \beta(2\nu_j - a) = 0.$$

Since

$$F''(\nu_j) = \frac{1}{\nu_j} + \frac{1}{a - \nu_j} - 2\beta > 0,$$

$F'(\nu_j)$ is strictly increasing from negative values for all ν_j near 0 to positive values for all ν_j near a. It follows that the only root of $F'(\nu_j) = 0$ is $\nu_j = a/2$ and thus that $\nu_k = a/2 = \nu_j$. Being a global minimizer of $R(\nu|\rho) - (\beta/2)\langle \nu, \nu\rangle$ over $\mathcal{P}$, $\overline{\nu}$ is also a global minimizer of the reduced two-variable problem. Since $a \in (0,1)$ is arbitrary, it follows that for any distinct pair of indices $\overline{\nu_j} = \overline{\nu_k}$. This completes the proof.

[1] Balian, R., *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*, Vol. I, translated by D. ter Haar and J. F. Gregg (Springer-Verlag, Berlin, 1991).
[2] Barré, J., Mukamel, D., and Ruffo, S., "Inequivalence of ensembles in a system with long-range interactions," Phys. Rev. Lett. **87**, 030601 (2001).
[3] Barré, J., Mukamel, D., and Ruffo, S., "Ensemble inequivalence in mean-field models of magnetism," in *Dynamics and Thermodynamics of Systems with Long Range Interactions*, edited by T. Dauxois, S. Ruffo, E. Arimondo, and M. Wilkens, Lect. Notes Phys., Vol. 602 (Springer-Verlag, New York, 2002), pp. 45–67.
[4] Biskup, M., and Chayes, L., "Rigorous analysis of discontinuous phase transitions via mean-field bounds," Commun. Math. Phys. **238**, 53–93 (2003).
[5] Borges, E. P., and Tsallis, C., "Negative specific heat in a Lennard-Jones-like gas with long-range interactions," Physica A **305**, 148–151 (2002).
[6] Caglioti, E., Lions, P. L., Marchioro, C., and Pulvirenti, A., "A special class of stationary flows for two dimensional Euler equations: a statistical mechanics description," Commun. Math. Phys. **143**, 501–525 (1992).
[7] Challa, M. S. S., and Hetherington, J. H., "Gaussian ensemble: an alternate Monte-Carlo scheme," Phys. Rev. A **38**, 6324–6337 (1988).
[8] Challa, M. S. S., and Hetherington, J. H., "Gaussian ensemble as an interpolating ensemble," Phys. Rev. Lett. **60**, 77–80 (1988).
[9] Costeniuc, M., Ellis, R. S., and Touchette, H., The Gaussian ensemble and universal ensemble equivalence for the Curie-Weiss-Potts model (unpublished).
[10] Costeniuc, M., Ellis, R. S., Touchette, H., and Turkington, B., "The generalized canonical ensemble and its universal equivalence with the microcanonical ensemble," cond-mat/0408681, J. Stat. Phys. (2005), to appear.
[11] Dauxois, T., Holdsworth, P., and Ruffo, S., "Violation of ensemble equivalence in the antiferromagnetic mean-field XY model," Eur. Phys. J. B **16**, 659 (2000).

[12] Dauxois, T., Latora, V., Rapisarda, A., Ruffo, S., and Torcini, A., "The Hamiltonian mean field model: from dynamics to statistical mechanics and back," in *Dynamics and Thermodynamics of Systems with Long-Range Interactions*, edited by T. Dauxois, S. Ruffo, E. Arimondo, and M. Wilkens, Lect. Notes Phys., Vol. 602 (Springer-Verlag, New York, 2002), pp. 458–487.
[13] Eisele, T., and Ellis, R. S., "Symmetry breaking and random waves for magnetic systems on a circle," Z. Wahrscheinlichkeitstheor. Verwandte Geb. **63**, 297–348 (1983).
[14] Ellis, R. S., *Entropy. Large Deviations and Statistical Mechanics* (Springer-Verlag, New York, 1985).
[15] Ellis, R. S., Haven, K., and Turkington, B., "Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles," J. Stat. Phys. **101**, 999–1064 (2000).
[16] Ellis, R. S., Haven, K., and Turkington, B., "Nonequivalent statistical equilibrium ensembles and refined stability theorems for most probable flows," Nonlinearity **15**, 239–255 (2002).
[17] Ellis, R. S., Otto, P., and Touchette, H., "Analysis of phase transitions in the mean-field Blume–Emery–Griffiths model," cond-mat/0409047, Ann. Appl. Prob. (2005), to appear.
[18] Ellis, R. S., Touchette, H., and Turkington, B., "Thermodynamic versus statistical nonequivalence of ensembles for the mean-field Blume–Emery–Griffiths model," Physica A **335**, 518–538 (2004).
[19] Ellis, R. S., and Wang, K., "Limit theorems for the empirical vector of the Curie–Weiss–Potts model," Stochastic Proc. Appl. **35**, 59–79 (1990).
[20] Ellis, R. S., and Wang, K., "Limit theorems for maximum likelihood estimators in the Curie–Weiss–Potts model," Stochastic Proc. Appl. **40**, 251–288 (1992).
[21] Eyink, G. L., and Spohn, H., "Negative-temperature states and large-scale, long-lived vortices in two-dimensional turbulence," J. Stat. Phys. **70**, 833–886 (1993).
[22] Gibbs, J. W., *Elementary Principles in Statistical Mechanics with Special Reference to the Rational Foundation of Thermodynamics* (Yale University Press, New Haven, 1902). Reprinted by Dover, New York 1960.
[23] Gross, D. H. E., "Microcanonical thermodynamics and statistical fragmentation of dissipative systems: the topological structure of the *n*-body phase space," Phys. Rep. **279**, 119–202 (1997).
[24] Gross, D. H. E., "Phase transitions without thermodynamic limit," in Proceedings of Les Houches Workshop on Nuclear Matter in Different Phases and Transitions, Les Houches, France, 31.3-10.4.98, edited by X. Campi, J. P. Blaizot, and M. Ploszaiczak (Kluwer Academic, New York, 1999), pp. 31–42.
[25] Hertel, P., and Thirring, W., "A soluble model for a system with negative specific heat," Ann. Phys. (N.Y.) **63**, 520 (1971).
[26] Hetherington, J. H., "Solid ^{3}He magnetism in the classical approximation," J. Low Temp. Phys. **66**, 145–154 (1987).
[27] Hetherington, J. H., and Stump, D. R., "Sampling a Gaussian energy distribution to study phase transitions of the Z(2) and U(1) lattice gauge theories," Phys. Rev. D **35**, 1972–1978 (1987).
[28] Huang, K., *Statistical Physics* (Wiley, New York 1987).
[29] Ispolatov, I., and Cohen, E. G. D., "On first-order phase transitions in microcanonical and canonical non-extensive systems," Physica A **295**, 475–487 (2000).
[30] Johal, R. S., Planes, A., and Vives, E., "Statistical mechanics in the extended Gaussian ensemble," Phys. Rev. E **68**, 056113 (2003).
[31] Kesten, H., and Schonmann, R. H., "Behavior in large dimensions of the Potts and Heisenberg models," Rev. Math. Phys. **1**, 147–182 (1990).
[32] Kiessling, M. K.-H., "On the equilibrium statistical mechanics of isothermal classical self-gravitating matter," J. Stat. Phys. **55**, 203–257 (1989).
[33] Kiessling, M. K.-H., and Lebowitz, J. L., "The micro-canonical point vortex ensemble: beyond equivalence," Lett. Math. Phys. **42**, 43–56 (1997).
[34] Kiessling, M. K.-H., and Neukirch, T., "Negative specific heat of a magnetically self-confined plasma torus," Proc. Natl. Acad. Sci. U.S.A. **100**, 1510–1514 (2003).
[35] Landau, L. D., and Lifshitz, E. M., *Statistical Physics*, 3rd ed., Vol. 5 of Landau and Lifshitz Course of Theoretical Physics (Butterworth Heinemann, Oxford, 1991).
[36] Latora, V., Rapisarda, A., and Tsallis, C., "Non-Gaussian equilibrium in a long-range Hamiltonian system," Phys. Rev. E **64**, 056134 (2001).
[37] Lynden-Bell, D., and Wood, R., "The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems," Mon. Not. R. Astron. Soc. **138**, 495 (1968).
[38] Messer, J., and Spohn, H., "Statistical mechanics of the isothermal Lane-Emden equation," J. Stat. Phys. **29**, 561–578 (1982).
[39] Pearce, P. A., and Griffiths, R. B., "Potts model in the many-component limit," J. Phys. A **13**, 2143–2148 (1980).
[40] Potts, R. B., "Some generalized order-disorder transformations," Proc. Cambridge Philos. Soc. **48**, 106–109 (1952).
[41] Reif, F., *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).
[42] Robert, R., and Sommeria, J., "Statistical equilibrium states for two-dimensional flows," J. Fluid Mech. **229**, 291–310 (1991).
[43] Rockafellar, R. T., *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).
[44] Salmon, R., *Lectures on Geophysical Fluid Dynamics* (Oxford University Press, New York, 1998).
[45] Smith, R. A., and O'Neil, T. M., "Nonaxisymmetric thermal equilibria of a cylindrically bounded guiding center plasma or discrete vortex system," Phys. Fluids B **2**, 2961–2975 (1990).
[46] Stump, D. R., and Hetherington, J. H., "Remarks on the use of a microcanonical ensemble to study phase transitions in the lattice gauge theory," Phys. Lett. B **188**, 359–363 (1987).
[47] Thirring, W., "Systems with negative specific heat," Z. Phys. **235**, 339–352 (1970).
[48] Touchette, H., Ellis, R. S., and Turkington, B., "An introduction to the thermodynamic and macrostate levels of non-

equivalent ensembles," Physica A **340**, 138–146 (2004).
[49] Wu, F. Y., "The Potts model," Rev. Mod. Phys. **54**, 235–268 (1982).