Journal of Statistical Physics, Vol. 77, Nos. 1/2, 1994

The Equivalence of Ensembles for Lattice Systems: Some Examples and a Counterexample

J. T. Lewis,[1] C.-E. Pfister,[2] and W. G. Sullivan[1,3]

Received December 31, 1993; final February 10, 1994

We describe the problem of the equivalence of ensembles at the level of states for classical lattice systems. We discuss circumstances where the vanishing of the specific information gain of a sequence of microcanonical measures with respect to a sequence of grand canonical measures implies the equivalence of ensembles. We give a simple derivation of a criterion for the vanishing of the specific information gain in terms of thermodynamic functions. The proof uses ideas from the theory of large deviations but is self-contained. We show how the criterion works in a simple model of a paramagnet and in the Ising model of a ferromagnet in any dimension but fails in the case of the Curie–Weiss mean-field model.

KEY WORDS: Microcanonical; grand canonical; equivalence of ensembles; large deviations; Ruelle–Lanford function; thermodynamic functions; specific information gain; entropy; paramagnet; Ising ferromagnet; Curie–Weiss mean-field model.

1. INTRODUCTION

Mark Kac was fond of saying that no theory is better than its best example. Oliver Penrose would surely echo this sentiment. He asked us for some examples to illustrate a theorem on the equivalence of ensembles[19]; we offer them to him, a pioneer of rigorous results in statistical mechanics who has never lost touch with the physical roots of the subject, on the occasion of his 65th birthday.

[1] Dublin Institute for Advanced Studies, Dublin 4, Ireland.

[2] Département de Mathématiques, CH-1015 Lausanne Switzerland.

[3] Department of Mathematics, University College, Belfield, Dublin 4, Ireland.

0022-4715/94/1000-0397$07.00/0 © 1994 Plenum Publishing Corporation

The concept of the equivalence of ensembles goes back to Gibbs,[11] whose development of statistical mechanics as 'the rational foundation of thermodynamics' is based on the canonical distribution. In Chapters X and XIV, he discusses the microcanonical distribution:

> From a certain point of view, the microcanonical distribution may seem more simple than the canonical, and it has perhaps been more studied, and been regarded as more closely related to the fundamental notions of thermodynamics. To this last point we shall return in a subsequent chapter. It is sufficient here to remark that analytically the canonical distribution is much more manageable than the microcanonical. (Ref. 11, Chapter X, p. 116.)
>
> As a general theorem, the conclusion may be expressed in the words: If a system of a great number of degrees of freedom is microcanonically distributed in phase, any very small part of it may be regarded as canonically distributed. (Ref. 11, Chapter XIV, p. 183.)

Since Gibbs' time, many proofs have been offered of this 'general theorem.' Not surprisingly, the case of noninteracting particles has received the greatest attention: Khinchine[16] in 1943 used a local central limit theorem to prove it for a classical ideal gas; a 'bare-hands' version of Khinchine's argument, using Stirling's formula, can be found in Martin-Löf.[22] There were many attempts to extend this to interacting systems using extensions of the central limit theorem to sums of weakly dependent random variables; see Mazur and van der Linden,[24] for example. In 1977, Dobrushin and Tirozzi[7] proved that the local central limit theorem is a consequence of the integral central limit theorem in the case of a Gibbs random field corresponding to a finite-range potential; however, their application of it to prove the equivalence of ensembles runs into problems when there is a first-order phase transition.

Typically, local central limit theorems hold on the scale of the square root of the volume. The right scale for the investigation of the equivalence of ensembles, however, turns out to be that of the volume itself; this is the scale on which large-deviation principles hold. Deuschel *et al.*[4] and Georgii[8] recently used a large-deviation principle for empirical measures to prove the equivalence of ensembles. One drawback with this approach is that it is technically difficult: since it involves measures on a space of measures, there are subtle points to be settled. Another is that the connection with thermodynamic functions is obscured.

A more elementary and direct approach[19] goes back to the common origin of large-deviation theory and statistical mechanics, the Principle of the Largest Term, and proves a result about the specific information gain of a sequence of conditioned measures (microcanonical ensembles) with respect to a sequence of tilted measures (grand canonical ensembles). This is a "soft" theorem—it uses nothing deeper than the order-completeness of the reals, but it is widely applicable. For noninteracting systems, the equivalence of ensembles for measures then follows from the Kemperman–Pinsker inequality[15,25] relating the information gain $\mathcal{H}(\mu \mid \nu)$ of μ with respect to ν to the total variation norm $\|\cdot\|_{\mathrm{TV}}$ of the difference of the two measures:

$$2\mathcal{H}(\mu \mid \nu) \geqslant \|\mu - \nu\|_{\mathrm{TV}}^2 \tag{1.1}$$

For interacting systems, our "soft" theorem has to be supplemented by a "hard" theorem, proved using the combinatorial devices introduced in the early 1970s by Sullivan[30] and perfected by Preston[26]; using it together with the Kemperman–Pinsker inequality, we prove that the vanishing of the specific information gain implies the equivalence of ensembles for a lattice gas with translational-invariant summable potentials.

Csiszár[3] seems to have been the first to use the vanishing of the specific information gain to prove a conditional limit theorem and this was the starting point of Georgii's work referred to above. What is new about our work is that we give a simple condition, stated in terms of thermodynamic functions, which is sufficient to ensure the vanishing of the specific information gain. Roughly stated, our main result is: *in the classical lattice gas, equivalence of ensembles at the level of states holds whenever it holds at the level of thermodynamic functions.* The aim of this paper is to explain our condition for the vanishing of the specific information gain and to illustrate it with some simple examples and a counterexample. These are elementary: the paramagnet, and the Ising model in one and two dimensions provide the examples; the counterexample, where equivalence of ensembles fails at the level of thermodynamic functions, is provided by the Curie–Weiss model.

In this Introduction, we have reviewed a small subset of the large body of work on the equivalence of ensembles, selected so as to set our work in context. To redress the balance a little, we mention another approach which is to be found in the works of Aizenmann *et al.*,[1] Georgii,[9] and Preston[27] and which is closer in spirit, perhaps, to Gibbs' original argument than proofs which make use of probabilistic reasoning; in this approach one works directly with infinite-volume states, defining the microcanonical and canonical states by local specifications as in the theory of Gibbs states. In contrast with the approach based on large deviations, this method runs into difficulties in the case of classical continuous-spin systems. We mention also the approach sketched by Lanford[17] and developed by Martin-Löf,[23] which is based on the Ruelle–Lanford function (see Section 4 for its definition), but which is less probabilistic in character than the proof we are about to describe, the main tools being convexity theory and the variational characterization of equilibrium states.

Here is the plan of the paper: In Section 2 we describe the models. In Section 3 we discuss the information gain and its use in proving the equivalence of ensembles and we begin the derivation of our criterion for the vanishing of the specific information gain. The expression we obtain for the specific information gain consists of three terms; the asymptotic behavior of these is treated in the three succeeding sections: in Section 4 we define the Ruelle–Lanford function related to the thermodynamic entropy; in Section 5 we discuss the concentration of measures; we give a proof of Varadhan's Theorem for compact spaces in Section 6. The treatment of all three topics relies on an abstract version of the Principle of the Largest Term, Lemma 4.1. Our criterion for the vanishing of the specific information gain is stated in Theorem 6.2; its application to the models is the subject of Section 7. In Section 8 we summarize our results.

2. THE MODELS

First, we set the notation: let $\mathbb{Z}^d$ $(d \geqslant 1)$ be an integer lattice, let $\{\Lambda_n\}_{n \geqslant 1}$ be an increasing sequence of cubes in $\mathbb{Z}^d$ with $V_n := |\Lambda_n| \to \infty$ as $n \to \infty$; at each site $j \in \Lambda_n$ we have a configuration space S_j which is a copy of the two-point set $S = \{-1, +1\}$. For each $n \geqslant 1$, the configuration space Ω_n is the space $\Omega_n = \prod_{j \in \Lambda_n} S_j$ which we regard as a subspace of the product space $\Omega = \prod_{j \in \mathbb{Z}^d} S_j$; equipped with the product topology, the space Ω is compact. The elementary random variables $\xi_j, j \in \mathbb{Z}^d$, are defined by

$$\xi_j(\omega) = \omega(j) \tag{2.1}$$

They have the following interpretation in the models: $\xi_j(\omega)$ is the magnetic moment at site j in the configuration ω. We define the *magnetization* M_n in Λ_n by

$$M_n = \sum_{j \in \Lambda_n} \xi_j \tag{2.2}$$

and the *pair interaction energy* U_n in Λ_n by

$$U_n = -\sum_{\langle i, j \rangle \subset \Lambda_n} \xi_i \xi_j \tag{2.3}$$

where $\langle i, j \rangle$ denotes a pair of nearest neighbor sites. We define the following models:

1. *The Paramagnet:* Let C be an open subinterval of $X = [-1, +1]$. Let ν_n^C be the microcanonical state obtained by conditioning on

M_n/V_n, the magnetization per site, taking values in C: for Σ a subset of Ω_n, we put

$$\nu_n^C[\Sigma] = \#\{\omega \in \Sigma: M_n(\omega)/V_n \in C\}/\#\{\omega \in \Omega_n: M_n(\omega)/V_n \in C\} \quad (2.4)$$

Let γ_n^t be the canonical state obtained by using the Boltzmann factor $\exp[tM_n(\omega)]$: for Σ a subset of Ω_n, we put

$$\gamma_n^t[\Sigma] = \sum_{\omega \in \Sigma} \exp[tM_n(\omega)] \Big/ \sum_{\omega \in \Omega_n} \exp[tM_n(\omega)] \quad (2.5)$$

2. *The Ising Model:* Let C be a convex open subset of $X = [-d, d] \times [-1, 1]$. Let ν_n^C be the microcanonical state obtained by conditioning on the pair $(U_n/V_n, M_n/V_n)$ taking values in C: for Σ a subset of Ω_n, we put

$$\nu_n^C[\Sigma] = \frac{\#\{\omega \in \Sigma: (U_n(\omega)/V_n, M_n(\omega)/V_n) \in C\}}{\#\{\omega \in \Omega_n: (U_n(\omega)/V_n, M_n(\omega)/V_n) \in C\}} \quad (2.6)$$

Let γ_n^t be the canonical state obtained by using the Boltzmann factor $\exp[t_1 U_n(\omega) + t_2 M_n(\omega)]$, where $t = (t_1, t_2)$: for Σ a subset of Ω_n, we put

$$\gamma_n^t[\Sigma] = \frac{\sum_{\omega \in \Sigma} \exp[t_1 U_n(\omega) + t_2 M_n(\omega)]}{\sum_{\omega \in \Omega_n} \exp[t_1 U_n(\omega) + t_2 M_n(\omega)]} \quad (2.7)$$

3. *The Curie–Weiss Model:* Let C be an open subinterval of $X = [-1, +1]$. Let ν_n^C be the microcanonical state obtained by conditioning on M_n/V_n, the magnetization per site, taking values in C: for Σ a subset of Ω_n, we put

$$\nu_n^C[\Sigma] = \frac{\sum_{\{\omega \in \Sigma: M_n(\omega)/V_n \in C\}} \exp[aM_n(\omega)^2/2V_n]}{\sum_{\{\omega \in \Omega_n: M_n(\omega)/V_n \in C\}} \exp[aM_n(\omega)^2/2V_n]} \quad (2.8)$$

where a is a positive real number. Let γ_n^t be the canonical state obtained by using the Boltzmann factor $\exp[tM_n(\omega)]$: for Σ a subset of Ω_n, we put

$$\gamma_n^t[\Sigma] = \frac{\sum_{\omega \in \Sigma} \exp[tM_n(\omega) + aM_n(\omega)^2/2V_n]}{\sum_{\omega \in \Omega_n} \exp[tM_n(\omega) + aM_n(\omega)^2/2V_n]} \quad (2.9)$$

The three models have certain features in common which it is worth abstracting:

- For $n \geqslant 1$, we have a *reference measure* ρ_n defined on the subsets of the finite set Ω_n: in the first two models, we have

$$\rho_n[\Sigma] = \#\{\omega \in \Sigma\} \quad (2.10)$$

while in the third we have

$$\rho_n[\Sigma] = \sum_{\omega \in \Sigma} \exp[aM_n(\omega)^2/2V_n] \tag{2.11}$$

• We have a *scale*, a sequence $V_0 := \{V_n\}_{n \geq 1}$ of positive numbers diverging to $+\infty$ as $n \to \infty$; in all three models we take $V_n = |\Lambda_n|$, the number of lattice sites in Λ_n.

• In each example, the microcanonical states are obtained by conditioning the reference measures; to do this, we use functions $T_n: \Omega_n \to X$ taking values in a compact subset X of a Euclidean space $E = \mathbb{R}^k$ $(k \geq 1)$: in the first and third examples,

$$T_n(\omega) := M_n(\omega)/V_n \tag{2.12}$$

$X = [-1, 1]$ and $k = 1$; in the second example,

$$T_n(\omega) := (U_n(\omega)/V_n, M_n(\omega)/V_n) \tag{2.13}$$

$X = [-d, d] \times [-1, 1]$ and $k = 2$. Conditioning is achieved by choosing a convex open subset C of X and restricting to those configurations ω for which $T_n(\omega) \in C$; the set

$$\{\omega \in \Omega_n : T_n(\omega) \in C\} \tag{2.14}$$

is sometimes called an *energy shell* in Ω_n.

• Both the microcanonical and the canonical states have *densities* with respect to the reference measures: the microcanonical state ν_n^C can be written with the aid of the function

$$\alpha_n^C(\omega) = 1_{T_n^{-1}C}(\omega)/\rho_n[T_n^{-1}C] \tag{2.15}$$

as

$$\nu_n^C[d\omega] = \alpha_n^C(\omega)\, \rho_n[d\omega] \tag{2.16}$$

The canonical state γ_n^t can be written with the aid of the function

$$\alpha_n^t(\omega) = \exp(V_n\langle t, T_n(\omega)\rangle) \Big/ \int_{\Omega_n} \exp(V_n\langle t, T_n(\omega)\rangle)\, \rho_n[d\omega] \tag{2.17}$$

as

$$\gamma_n^t[d\omega] = \alpha_n^t(\omega)\, \rho_n[d\omega] \tag{2.18}$$

Here we have made use of the inner product $\langle \cdot, \cdot \rangle$ defined on X by

$$\langle t, x\rangle := t_1 x_1 + \cdots + t_k x_k \tag{2.19}$$

3. THE INFORMATION GAIN

We recall the definition of the information gain (also known as the relative entropy, a name we avoid because 'entropy' has become overworked).

Definition 3.1. Let λ_1 and λ_2 be probability measures on a space Ω; the *information gain* $\mathscr{H}(\lambda_1 \mid \lambda_2)$ of λ_1 with respect to λ_2 is given by

$$\mathscr{H}(\lambda_1 \mid \lambda_2) := \begin{cases} \int_\Omega \ln h(\omega)\, \lambda_1[d\omega], & \text{if} \quad \lambda_1[d\omega] = h(\omega)\, \lambda_2[d\omega] \\ +\infty, & \text{otherwise} \end{cases} \tag{3.1}$$

Using (2.16) and (2.18), we have

$$\mathscr{H}(\nu_n^C \mid \gamma_n') = \sum_{\omega \in \Omega_n} [\ln \alpha_n^C(\omega) - \ln \alpha_n'(\omega)]\, \alpha_n^C(\omega)\, \rho_n[\omega] \tag{3.2}$$

where the integral in (3.1) becomes a sum because the reference measure ρ_n is discrete. In this case, we have

$$\|\nu_n^C - \gamma_n'\|_{\mathrm{TV}} = \sum_{\omega \in \Omega_n} |\alpha_n^C(\omega) - \alpha_n'(\omega)|\, \rho_n[\omega] \tag{3.3}$$

The total-variation distance between two measures is related to the information gain by the Kemperman–Pinsker inequality[15,25]:

$$2\mathscr{H}(\mu \mid \nu) \geqslant \|\mu - \nu\|_{\mathrm{TV}}^2 \tag{3.4}$$

This inequality is the key to using the information gain to prove the equivalence of ensembles. This is seen most clearly when the canonical state γ_n' is a product state; this is the case in our first example, the paramagnet.

We consider $\nu_{n,\Delta}^C$, the restriction to a subset Δ of $\Lambda_n \subset \mathbb{Z}^d$ of the microcanonical state ν_n^C. For the paramagnet, the canonical state γ_n' is a product measure; this has two important consequences:

1. The restriction of γ_n' to $\Delta \subset \Lambda_n$ is independent of n and we denote it by γ_Δ'.

2. If Δ_1 and Δ_2 are disjoint copies of Δ such that $\Delta_1 \cup \Delta_2 \subset \Lambda_n$, then

$$\mathscr{H}(\nu_{n,\Delta_1 \cup \Delta_2}^C \mid \gamma_{\Delta_1 \cup \Delta_2}') \geqslant \mathscr{H}(\nu_{n,\Delta_1}^C \mid \gamma_{\Delta_1}') + \mathscr{H}(\nu_{n,\Delta_2}^C \mid \gamma_{\Delta_2}') \tag{3.5}$$

but

$$\mathscr{H}(\nu_{n,\Delta_1}^C \mid \gamma_{\Delta_1}') = \mathscr{H}(\nu_{n,\Delta_2}^C \mid \gamma_{\Delta_2}') \tag{3.6}$$

so that

$$\mathscr{H}(\nu_n^C \mid \gamma_n') \geqslant \lfloor V_n / |\Delta| \rfloor\, \mathscr{H}(\nu_{n,\Delta}^C \mid \gamma_\Delta') \tag{3.7}$$

where $\lfloor x \rfloor$ denotes the integer part of x. Hence

$$\lim_{n \to \infty} \frac{1}{V_n} \mathscr{H}(\nu_n^C \mid \gamma_n') = 0 \tag{3.8}$$

implies that

$$\lim_{n \to \infty} \mathscr{H}(\nu_{n,\Delta}^C \mid \gamma_\Delta') = 0 \tag{3.9}$$

By (3.4), this implies that

$$\|\nu_{n,\Delta}^C - \gamma_\Delta'\|_{\text{TV}} \to 0 \tag{3.10}$$

and, by (3.3), the density $\alpha_{n,\Delta}^C$ of the restriction of the microcanonical state to the finite set Δ converges to the density α_Δ' of the canonical state associated with the set Δ. So, for the paramagnet, Gibbs' 'general theorem' holds provided we can prove (3.8), the vanishing of the specific information gain.

In our second example, the Ising model, the canonical state γ_n' is not a product measure, because of the interaction between spins on neighboring sites and the subadditivity argument fails; this can be got around, using the techniques of Sullivan[30] and Preston.[26] There is a second difficulty in the case of the Ising model: at (3.6), we exploited permutation invariance (exchangeability); this no longer holds and we must exploit the translation invariance instead. However, the microcanonical state ν_n^C associated with the finite box Λ_n is not translation-invariant; the way out is to replace it by its spatial average. For each $j \in \mathbb{Z}^d$ we have the action of $\mathbb{Z}^d$ on itself given by $i \to i + j$, $i \in \mathbb{Z}^d$; this lifts to $\theta_j: \Omega \to \Omega$ given by $(\theta_j \omega)(i) = \omega(i - j)$ for each configuration $\omega \in \Omega$. We choose some fixed configuration $\eta \in \Omega$ and define a map $b_n^\eta: \Omega_n \to \Omega$ by $b_n^\eta(\omega) := \eta(i)$ if $i \notin \Lambda_n$ and $b_n^\eta(\omega) := \omega(i)$ if $i \in \Lambda_n$; the image measure $\nu_n^{C,\eta} = \nu_n^C \circ (b_n^\eta)^{-1}$ is defined on Ω and we define the averaged microcanonical state $\bar{\nu}_n^{C,\eta}$ by

$$\bar{\nu}_n^{C,\eta} := \frac{1}{V_n} \sum_{j \in \Lambda_n} \nu_n^{C,\eta} \circ \theta_j^{-1} \tag{3.11}$$

[If in (2.3) we sum over all pairs $\langle i, j \rangle$ of nearest neighbor sites which have a nonempty intersection with the cube Λ_n, then the measure $\nu_n^{C,\eta}$ is the usual Gibbs measure in Λ_n with boundary condition η; our results apply to this case.] For the Ising model and, indeed, for any model with translation-invariant summable potentials, we are able to prove the following.

Condition (3.8), the vanishing of the specific information gain, implies that any weak limit point of the sequence $\{\bar{\nu}_n^{C,n}\}_{n\geq 1}$ is a Gibbs state with respect to the specification associated with $\{\gamma_n^t\}_{n\geq 1}$.

This is our version of the 'general theorem' of Gibbs in the interacting case; a full proof will be given elsewhere.[20] Here we concentrate on obtaining a criterion to determine when, given a sequence $\{\nu_n^C\}_{n\geq 1}$ of microcanonical states, we can choose t so that the specific information gain

$$\lim_{n\to\infty} \frac{1}{V_n} \mathscr{H}(\nu_n^C \mid \gamma_n^t) \tag{3.12}$$

vanishes; we will illustrate the use of our criterion with our three examples.

The key to the derivation of our criterion is the following observation: *the density α_n^C of the microcanonical state and the density α_n^t of the canonical state are both functions of T_n.* This enables us to express the information gain as an integral over the space X by using the change-of-variables formula: let $T: \Omega \to X$ be an X-valued random variable and let $\mathbb{K} = \mathbb{P} \circ T^{-1}$ be the image law of $\mathbb{P}$ under T defined by $\mathbb{K}[B] = \mathbb{P}[\{\omega: T(\omega) \in B\}]$; the random variable $f: X \to \mathbb{R}$ is $\mathbb{K}$-integrable if and only if $f \circ T$ is $\mathbb{P}$-integrable and then

$$\int_\Omega (f \circ T)(\omega)\, \mathbb{P}[d\omega] = \int_X f(x)\, \mathbb{K}[dx] \tag{3.13}$$

Define the distribution $\mathbb{M}_n$ of T_n under ρ_n by $\mathbb{M}_n := \rho_n \circ T_n^{-1}$; we have

$$\nu_n^C \circ T_n^{-1} = \frac{\mathbb{M}_n[\cdot \cap C]}{\mathbb{M}_n[C]} =: \mathbb{M}_n[\cdot \mid C] \tag{3.14}$$

$$\gamma_n^t \circ T_n^{-1} = \frac{\mathbb{M}_n^t[\cdot]}{\mathbb{M}_n^t[X]} =: \mathbb{M}_n^t[\cdot \mid X] \tag{3.15}$$

where $\mathbb{M}_n^t[dx] := \exp(V_n\langle t, x\rangle)\, \mathbb{M}_n[dx]$. Thus, using the change-of-variable formula, we have

$$\mathscr{H}(\nu_n^C \mid \gamma_n^t) = \mathscr{H}(\mathbb{M}_n[\cdot \mid C] \mid \mathbb{M}_n^t[\cdot \mid X]) \tag{3.16}$$

We see that the density of $\mathbb{M}_n[\cdot \mid C]$ with respect to $\mathbb{M}_n$ is $1_C(\cdot)/\mathbb{M}_n[C]$ and the density of $\mathbb{M}_n^t[\cdot \mid X]$ with respect to $\mathbb{M}_n$ is $\exp(V_n\langle t, \cdot\rangle)/\exp(V_n p_n(t))$, where $p_n(t)$ is defined by

$$\exp(V_n p_n(t)) := \int_X \exp(V_n\langle t, x\rangle)\, \mathbb{M}_n[dx] \tag{3.17}$$

It follows, using (3.16), that

$$\frac{1}{V_n} \mathscr{H}(\nu_n^C \mid \gamma_n^t) = -\int_C \langle t, x \rangle \, \mathbb{M}_n[dx \mid C] + p_n(t) - \frac{1}{V_n} \ln \mathbb{M}_n[C] \tag{3.18}$$

It remains to evaluate the limits as $n \to \infty$ of each of the three terms on the right-hand side of (3.18). To do this, it is necessary to review some results relating to the Ruelle–Lanford function[29,17]; proofs can be found in ref. 18.

4. THE RUELLE–LANFORD FUNCTION

We need to examine the behavior as $n \to \infty$ of the measures on X defined in Section 2. Since Ω_n, ρ_n, and T_n play no part in the considerations of this section, it is best to start afresh. Let $\mathbb{M}_0 := \{\mathbb{M}_n\}_{n \geq 1}$ be a sequence of finite positive measures on X, a compact convex subset of $E = \mathbb{R}^k$. Let V_0 be a scale; define set functions m_n, $\underline{m}$, $\bar{m}$ on subsets of X by

$$m_n[B] := \frac{1}{V_n} \ln \mathbb{M}_n[B] \tag{4.1}$$

$$\underline{m}[B] := \liminf_{n \to \infty} m_n[B] \tag{4.2}$$

$$\bar{m}[B] := \limsup_{n \to \infty} m_n[B] \tag{4.3}$$

The following properties are straightforward consequences of the definitions:

$$\underline{m}[B] \leqslant \bar{m}[B] \qquad \text{for all} \quad B \tag{4.4}$$

$$\underline{m} \quad \text{and} \quad \bar{m} \quad \text{are increasing} \tag{4.5}$$

The next property is an abstract version of the Principle of the Largest Term, well known in statistical mechanics (see, for example, Huang[13]). Since it is central to our development, we give a proof. (For a, $b \in \mathbb{R}$, we denote the maximum of a and b by $a \vee b$.)

Lemma 4.1. On $\mathscr{B}(X)$, we have

$$\bar{m}[B_1 \cup B_2] = \bar{m}[B_1] \vee \bar{m}[B_2] \tag{4.6}$$

Proof. For $j = 1, 2$, we have

$$\mathbb{M}_n[B_j] \leqslant \mathbb{M}_n[B_1 \cup B_2] \leqslant \mathbb{M}_n[B_1] + \mathbb{M}_n[B_2] \tag{4.7}$$

so that

$$\mathbb{M}_n[B_1] \vee \mathbb{M}_n[B_2] \leqslant \mathbb{M}_n[B_1 \cup B_2] \leqslant 2\mathbb{M}_n[B_1] \vee \mathbb{M}_n[B_2] \tag{4.8}$$

It follows that

$$\bar{m}[B_1 \cup B_2] = \limsup_{n\to\infty}(m_n[B_1] \vee m_n[B_2]) \tag{4.9}$$

But for each pair $\{a_n\}_{n\geqslant 1}$, $\{b_n\}_{n\geqslant 1}$ of sequences of real numbers, we have

$$\limsup_{n\to\infty}(a_n \vee b_n) = (\limsup_{n\to\infty} a_n) \vee (\limsup_{n\to\infty} b_n) \tag{4.10}$$

Thus (4.6) follows from (4.9) and (4.10). ■

Define functions $\underline{\mu}$, $\bar{\mu}$ on X as follows:

$$\underline{\mu}(x) := \inf\{\underline{m}[G]\colon G \text{ open}, G \ni x\} \tag{4.11}$$

$$\bar{\mu}(x) := \inf\{\bar{m}[G]\colon G \text{ open}, G \ni x\} \tag{4.12}$$

The following properties are direct consequences of the definitions:

$$\underline{\mu} \quad \text{and} \quad \bar{\mu} \quad \text{are upper semicontinuous functions} \tag{4.13}$$

$$\bar{m}[G] \geqslant \sup_{x\in G} \bar{\mu}(x), \qquad G \text{ open} \tag{4.14}$$

$$\underline{m}[G] \geqslant \sup_{x\in G} \underline{\mu}(x), \qquad G \text{ open} \tag{4.15}$$

The lower bound (4.14) for $\bar{m}$ on open sets is rarely used; of greater importance is the following upper bound for $\bar{m}$ on compact sets, a consequence of the Principle of the Largest Term (4.6):

$$\bar{m}[K] \leqslant \sup_{x\in K} \bar{\mu}(x), \qquad K \text{ compact} \tag{4.16}$$

If $\bar{\mu}(x) = \underline{\mu}(x)$ for all $x \in X$, we say the *Ruelle–Lanford function* (RL-function) μ *exists* for the pair $(\mathbb{M}_0, V_0)$ and is given by

$$\mu(x) := \underline{\mu}(x) = \bar{\mu}(x) \tag{4.17}$$

When the RL-function exists, the bounds (4.15) and (4.16) can be restated as

$$\bar{m}[K] \leqslant \sup_{x\in K} \mu(x), \qquad K \text{ compact} \tag{4.18}$$

$$\underline{m}[G] \geqslant \sup_{x\in G} \mu(x), \qquad G \text{ open} \tag{4.19}$$

When (4.18) and (4.19) hold, we say (following Varadhan[31] that *a large-deviation principle* (LDP) *holds with rate function* $I = -\mu$ *for the pair* $(\mathbb{M}_0, V_0)$.

This means that the sequence m_0 of set functions m_n, defined at (4.1), converges to the set function

$$B \mapsto \sup_{x \in B} \mu(x) \tag{4.20}$$

in *exactly the same sense* that a sequence of probability measures $\mathbb{M}_0$ converges to a measure δ_x in a weak law of large numbers (remember that X is assumed to be compact). For special sets, convergence takes a more familiar form: for example, *if X is compact, C is a nonempty open convex subset of X, and the RL-function μ exists and is concave, then*

$$\sup_{x \in C} \mu(x) = \underline{m}[C] = \bar{m}[C] = \sup_{x \in \bar{C}} \mu(x) \tag{4.21}$$

and the bounds (4.18) and (4.19) yield

$$\lim_{n \to \infty} m_n[C] = \sup_{x \in \bar{C}} \mu(x) \tag{4.22}$$

This result enables us to deal with the third term in (3.18).

Remarks.

- We have given μ the name "Ruelle–Lanford function" because, in the setting of a lattice gas with translation-invariant summable potentials, our definition coincides with the definition of entropy given by Ruelle[29] and Lanford.[17]

- We reserve the name 'entropy' for those RL-functions which are concave. Lanford's proof of the existence of the entropy in the case of the lattice gas yields also its concavity. The Curie–Weiss model provides an example of an RL-function which is not concave.

- Ruelle and Lanford understood that giving precise meaning to Boltzmann's formula

$$S = k \ln W \tag{4.23}$$

relating the entropy S of a macroscopic equilibrium state to the number W of corresponding microscopic states, is the *same problem* as that of making sense of the convergence of the sequence m_0 to the set function (4.20); by so doing, they introduced a new technique in the theory of large deviations (compare Bahadur and Zabel[2]).

In the models we are considering, the RL-function exists and the existence is established in a variety of ways:

1. *The Paramagnet:* The measure $\mathbb{M}_n^0$ is defined by

$$\mathbb{M}_n^0[B] := \#\{\omega \in \Omega_n : M_n(\omega)/V_n \in B\} \tag{4.24}$$

for $B \subset [-1, 1]$. Using Stirling's formula and some elementary analysis, we can prove that the RL-function μ^0 exists and is given by

$$\mu^0(x) = \ln 2 - \tfrac{1}{2}v(1-x) - \tfrac{1}{2}v(1+x) \tag{4.25}$$

where

$$v(x) := \begin{cases} 0, & x = 0 \\ x \ln x, & 0 < x < 1 \\ 0, & x = 1 \end{cases} \tag{4.26}$$

2a, *The Ising Model* ($d = 1$): The measure $\mathbb{M}_n^1$ is defined by

$$\mathbb{M}_n^1[B] := \#\{\omega \in \Omega_n : (U_n(\omega)/V_n, M_n(\omega)/V_n) \in B\} \tag{4.27}$$

for $B \subset [-1, 1] \times [-1, 1]$. Using the combinatorics in Ising's original treatment,[14] Stirling's formula, and some elementary analysis, we can prove that the RL-function μ^1 exists and is given by

$$\mu^1(x_1, x_2) = \begin{cases} \ln 2 - \tfrac{1}{2}v(1+x_1) + \tfrac{1}{2}v(1+x_2) + \tfrac{1}{2}v(1-x_2) & \\ \quad -\tfrac{1}{4}v(1-x_1-2x_2) - \tfrac{1}{4}v(1-x_1+2x_2), & x \in \Delta \\ -\infty, & x \notin \Delta \end{cases} \tag{4.28}$$

where

$$\Delta := \left\{ x \in X: -1 \leqslant x_1 \leqslant 1, -\frac{1-x_1}{2} \leqslant x_2 \leqslant \frac{1-x_1}{2} \right\} \tag{4.29}$$

The details of this calculation will be given in ref. 21; compare ref. 10.

2b. *The Ising Model* ($d \geqslant 2$): The measure $\mathbb{M}_n^d$ is defined by

$$\mathbb{M}_n^d[B] := \#\{\omega \in \Omega_n : (U_n(\omega)/V_n, M_n(\omega)/V_n) \in B\} \tag{4.30}$$

for $B \subset [-d, d] \times [-1, 1]$. For $d \geqslant 2$, we no longer have the benefit of explicit expressions for the RL-function and must rely on the general arguments which apply to the lattice gas with translation-invariant summable potentials. We sketch these here; details are given in ref. 20. Using standard methods, we prove that $\bar{\mu}$ and $\underline{\mu}$ are independent of boundary conditions. Let $B_\varepsilon(x)$ be an open ball of radius ε and center x in X; we prove, in the case of free boundary conditions, the following result:

Let x, x', $x'' \in X$ *satisfy* $x' + x'' = 2x$ *and let* $0 < \varepsilon' < \varepsilon$; *then*

$$2\underline{m}[B_\varepsilon(x)] \geqslant \underline{m}[B_{\varepsilon'}(x')] + \bar{m}[B_{\varepsilon'}(x'')] \tag{4.31}$$

From this and the independence of $\bar{\mu}$ and $\underline{\mu}$ on the boundary conditions, we deduce that:

The RL-function μ^d *exists for the pair* $(\mathbb{M}_0^d, V_0)$ *and is concave on* X.

3. *The Curie–Weiss Model:* The measure $\mathbb{M}_n^{\mathrm{CW}}$ is defined by

$$\mathbb{M}_n^{\mathrm{CW}}[dx] = \exp(V_n\, g(x))\, \mathbb{M}_n^0[dx] \tag{4.32}$$

where $g(x) := ax^2/2$. We shall prove in Section 6 that the RL-function μ^{CW} exists and is given by

$$\mu^{\mathrm{CW}}(x) = \mu^0(x) + ax^2/2 \tag{4.33}$$

Notice that $x \to \mu^{\mathrm{CW}}(x)$ is not concave for $a > 1/2$.

5. CONCENTRATION OF MEASURES

Our next application of (4.16) is to the *concentration of measures.* The results of this section will enable us to deal with the first term on the right-hand side of (3.18). Let $\mathbb{M}_0$ be a sequence of probability measures; if $\mathbb{M}_0$ converges weakly to a Dirac measure δ_x at some point $x \in X$, we say $\mathbb{M}_0$ obeys a weak law of large numbers (WLLN). In the absence of a first-order phase transition, a WLLN holds in the grand canonical ensemble. We require a substitute for a WLLN which holds regardless of phase transitions. We say that a sequence $\mathbb{M}_0$ of probability measures on X is *eventually concentrated on a set* A if, for each open neighborhood G of A, we have

$$\lim_{n \to \infty} \mathbb{M}_n[G] = 1 \tag{5.1}$$

We shall need the following result.

Lemma 5.1. Let $\mathbb{M}_0$ be a sequence of probability measures which is eventually concentrated on a set A; if $f\colon X \to \mathbb{R}$ is lower semicontinuous and bounded below on X, then

$$\inf_{x \in A} f(x) \leqslant \liminf_{n \to \infty} \int_X f(x)\, \mathbb{M}_n[dx] \tag{5.2}$$

[There is an obvious complementary upper bound; together they yield the usual characterization of the WLLN in terms of bounded continuous functions when A reduces to a single point: if $A = \{x\}$ and $\mathbb{M}_0$ is eventually concentrated on A, then $\mathbb{M}_0$ converges weakly to the Dirac measure δ_{x}.] The function $\bar{\mu}$, defined at (4.12) for the pair $(\mathbb{M}_0, V_0)$, enables us to determine a concentration set for the sequence $\mathbb{M}_0$. (How useful it is depends on how well we have chosen the scale V_0.) Notice that, for probability measures, the function $\bar{\mu}$ is bounded above by zero; in fact, it always attains this bound and the set on which it attains it is a concentration set for $\mathbb{M}_0$. Let $N_{\bar{\mu}}$ be the set defined by

$$N_{\bar{\mu}} := \{x \in X: \bar{\mu}(x) = 0\} \tag{5.3}$$

Lemma 5.2. Let $\mathbb{M}_0$ be a sequence of probability measures and V_0 a scale. Then:

(a) $N_{\bar{\mu}}$ is compact and nonempty;

(b) The sequence $\mathbb{M}_0$ is eventually concentrated on $N_{\bar{\mu}}$.

The proofs of both (a) and (b) make use of the bound (4.16).

We are now in a position to deal with the first term on the right-hand side of (3.18). Let C be an open convex subset of X; using convexity theory, we can prove that *if the RL-function μ exists for the pair $(\mathbb{M}_0, V_0)$ and is concave, then the RL-function μ_C for the pair $(\mathbb{M}_0[\cdot \mid C], V_0)$ exists and is given by*

$$\mu_C(x) = \begin{cases} \mu(x) - \sup_{y \in \bar{C}} \mu(y), & x \in \bar{C} \\ -\infty, & x \in X \setminus \bar{C} \end{cases} \tag{5.4}$$

Applying Lemma 5.2 to the sequence $\mathbb{M}_0[\cdot \mid C]$ of probability measures, we see that *the sequence $\mathbb{M}_0[\cdot \mid C]$ is eventually concentrated on the set*

$$X_{\bar{C}} := N_{\mu_C} = \{x \in X: \mu(x) = \sup_{y \in \bar{C}} \mu(y)\} \tag{5.5}$$

so that, by Lemma 5.1, we have

$$\liminf_{n \to \infty} \int_C \langle t, x \rangle \, \mathbb{M}_n[dx] \geqslant \inf_{X_C} \langle t, x \rangle \tag{5.6}$$

Remark. As already mentioned, when the RL-function is concave, μ is interpreted as the thermodynamic entropy; Lemma 5.2 and formula (5.5) give an expression of the Maximum Entropy Principle.

6. VARADHAN'S THEOREM

It remains to deal with the middle term on the right-hand side of (3.18) and to compute the RL-function for the Curie–Weiss model; for both of these, we need yet another consequence of the Principle of the Largest Term: Varadhan's Theorem in the case when the space X is compact. Let $g: X \to \mathbb{R}$ be continuous; define the measure $\mathbb{M}_n^g$ by $\mathbb{M}_n^g[dx] := \exp(V_n g(x))\, \mathbb{M}_n[dx]$; let $\bar{\mu}^g, \underline{\mu}^2$ be the upper and lower functions determined by the pair $(\mathbb{M}_0^g, V_0)$; they are related to $\bar{\mu}$ and $\underline{\mu}$ as follows:

$$\bar{\mu}^g(x) = \bar{\mu}(x) + g(x) \tag{6.1}$$

$$\underline{\mu}^g(x) = \underline{\mu}(x) + g(x) \tag{6.2}$$

These relations are a consequence of the continuity of the function $x \mapsto g(x)$. We are now ready for our third application of the bound (4.16):

Theorem 6.1. Suppose that the RL-function μ exists for the pair $(\mathbb{M}_0, V_0)$ and that the function $x \to g(x)$ is continuous; then:

(a) The RL-function μ^g exists for the pair $(\mathbb{M}_0^g, V_0)$

(b) The pair $(\mathbb{M}_0^g, V_0)$ obeys an LDP:

$$\bar{m}^g[K] \leqslant \sup_{x \in K} \mu^g(x), \qquad K \text{ compact} \tag{6.3}$$

$$\underline{m}^g[G] \geqslant \sup_{x \in G} \mu^g(x), \qquad G \text{ open} \tag{6.4}$$

(c) μ^g is given by

$$\mu^g(x) = g(x) + \mu(x) \tag{6.5}$$

(d) If, in addition, the space X is compact, then

$$\lim_{n \to \infty} \frac{1}{V_n} \ln \int_X \exp(V_n g(x))\, \mathbb{M}_n[dx] = \sup_{x \in X}(g(x) + \mu(x)) \tag{6.6}$$

Proof. Both (a) and (c) follow from (6.1) and (6.2); (b) follows from (a) and (4.18) and (4.19). To deduce (d) from (c), note that X is both compact and open as a topological space so that, from (6.3) and (6.4), we have

$$\begin{aligned} &\limsup_{n \to \infty} \frac{1}{V_n} \ln \int_X \exp(V_n g(x))\, \mathbb{M}_n[dx] \\ &\leqslant \sup_{x \in X} \mu^g(x) \\ &\leqslant \liminf_{n \to \infty} \frac{1}{V_n} \ln \int_X \exp(V_n g(x))\, \mathbb{M}_n[dx] \quad \blacksquare \end{aligned} \tag{6.7}$$

Our first application is to the Curie–Weiss model; it follows from the existence of the RL-function μ^0 for the pair $(\mathbb{M}_0^0, V_0)$ and the continuity of the function $x \to ax^2/2$ that the RL-function μ^{CW} exists for the pair $(\mathbb{M}_0^{CW}, V_0)$ and is given by

$$\mu^{CW}(x) = \mu^0(x) + ax^2/2 \tag{6.8}$$

Our second application is to the remaining term in (3.18); recall that $p_n(t)$ was defined at (3.17) as

$$p_n(t) := \frac{1}{V_n} \ln \int_X \exp(V_n\langle t, x\rangle)\, \mathbb{M}_n[dx] \tag{6.9}$$

Since $x \to \langle t, x\rangle$ is continuous, it follows from (d) that *if the RL-function* μ *exists, then the* grand canonical pressure *defined by*

$$p(t) := \lim_{n\to\infty} p_n(t) \tag{6.10}$$

exists and is given by

$$p(t) = \sup_{x\in X}(\langle t, x\rangle + \mu(x)) \tag{6.11}$$

We can write (6.11) as $p(t) = (-\mu)^*(t)$, where f^* is the Legendre transform of f:

$$f^*(y) := \sup_{x\in X}(\langle y, x\rangle - f(x)) \tag{6.12}$$

We are now ready to prove our criterion for the vanishing of the specific information gain. To state it, we define the set X^t for $t \in \mathbb{R}^k$ by

$$X^t := \{x \in X\colon p(t) = \langle t, x\rangle + \mu(x)\} \tag{6.13}$$

Theorem 6.2. Suppose that C is an open convex subset of the compact space X and that the RL-function μ exists for the pair $(\mathbb{M}_0, V_0)$ and that μ is finite at some point of C; if $X_C \subset X^t$, then the specific information gain is zero:

$$\lim_{n\to\infty} \frac{1}{V_n} \mathscr{H}(\nu_n^C \mid \gamma_n^t) = 0 \tag{6.14}$$

Proof. By (3.18), we have

$$\frac{1}{V_n}\mathscr{H}(\nu_n^C \mid \gamma_n^t) = \frac{1}{V_n}\mathscr{H}(\mathbb{M}_n[\cdot \mid C] \mid \mathbb{M}_n^t[\cdot \mid X])$$

$$= -\int \langle t, y\rangle\, \mathbb{M}_n[dy \mid C] + m_n^t[X] - m_n[C] \tag{6.15}$$

By (3.4), the left-hand side of (6.14) is nonnegative; by (5.6), (6.9), and (4.22), we have an upper bound:

$$0 \leqslant \limsup_{n \to \infty} \frac{1}{V_n} \mathscr{H}(v_n^C \mid \gamma_n') \leqslant -\inf_{y \in X_C} \langle t, y \rangle + p(t) - \sup_{y \in X_C} \mu(y)$$

$$= \sup_{y \in X_C} \{ p(t) - \langle t, y \rangle - \mu(y) \} \tag{6.16}$$

If $X_C \subset X'$, then

$$\sup_{y \in X_C} \{ p(t) - \langle t, y \rangle - \mu(y) \} = 0 \tag{6.17}$$

and (6.14) follows. ■

Remark. By (c) of Theorem 6.1, the RL-function of the pair $(\mathbb{M}_0'[\cdot \mid X], V_0)$ is $\mu(x) + \langle t, x \rangle - p(t)$, so that, by Lemma 5.2, the set X' is nonempty and is a concentration set for the sequence $\mathbb{M}_0'[\cdot \mid X]$ of probability measures. By (6.11), the grand canonical pressure p is the Legendre transform $(-\mu)^*$ of $-\mu$ and it is not difficult to show that X' is a subset of the *subdifferential* ∂p of p at t:

$$X' \subset \partial p(t) := \{ x : p(t+s) \geqslant p(t) + \langle s, x \rangle, \forall s \} \tag{6.18}$$

If, in addition, we have equivalence of ensembles at the level of thermodynamic functions in the sense that $\mu(x) = -p^*(x)$ as well as $p(t) = (-\mu)^*(t)$, which is the case when μ is concave, then we have equality: $X' = \partial p(t)$.

7. APPLICATIONS OF THE CRITERION

To illustrate how Theorem 6.2 may be applied, we consider in turn the three models described above.

1. *The paramagnet:* Choose $C = (c', c'') \subset [-1, 1]$; the RL-function μ^0 exists for the pair $(\mathbb{M}_0^0, V_0)$ and is given by (4.25) and the grand canonical pressure $p(t)$ is given by

$$p(t) = \ln 2 \cosh t \tag{7.1}$$

The set $X_C = \{x^*\}$, where

$$x^* = \begin{cases} c', & 0 \leqslant c' \\ 0, & c' < 0 < c'' \\ c'', & c'' \leqslant 0 \end{cases} \tag{7.2}$$

and the set $X^t = \{x_t\}$, where

$$x_t = p'(t) = \tanh t \tag{7.3}$$

Given C, we can find t^* such that $X_{\bar{C}} = X^{t^*}$; thus we have

$$\lim_{n \to \infty} \frac{1}{V_n} \mathscr{H}(v_n^C \mid \gamma_n^{t^*}) = 0 \tag{7.4}$$

2a. *The Ising Model* ($d = 1$): Choose $C = (c_1', c_1'') \times (c_2', c_2'')$; the RL-function μ^1 exists and is given by (4.28); since μ^1 is strictly concave on the set Δ on which it is finite, the set $X_{\bar{C}}$ consists of a single point x^* provided C contains at least one point at which μ^1 is finite. See Fig. 1.

Since the RL-function μ^1 exists, it follows from (6.11) that the grand canonical pressure $p^1(t)$ exists and is given by

$$p^1(t) = \sup_{x \in X}(\langle t, x \rangle + \mu^1(x)) \tag{7.5}$$

Since μ^1 is strictly concave on the set Δ on which it is finite, it follows from convexity theory (see ref. 28, for example) that p^1 is continuously differentiable on the set on which it is finite and the set X^t consists of the single point

$$x_t := \text{grad } p(t) \tag{7.6}$$

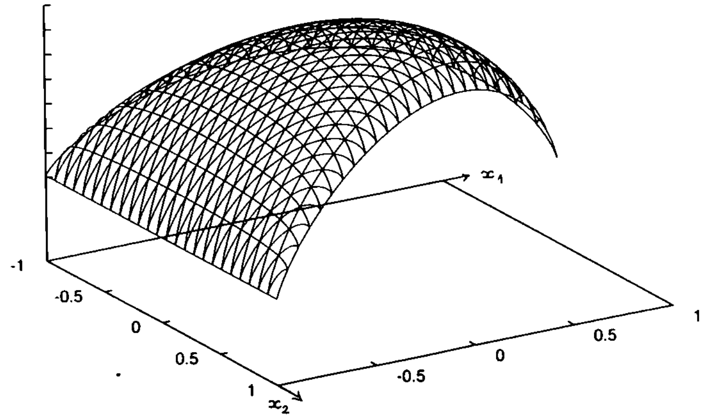

Fig. 1. The entropy surface of the Ising model ($d = 1$) plotted against pair interaction energy per site (x_1) and magnetization per site (x_2).

Choosing t^* to be the unique root of

$$\text{grad } p(t) = x^* \tag{7.7}$$

we have $X_{\mathcal{C}} = X^{t^*}$, so that

$$\lim_{n \to \infty} \frac{1}{V_n} \mathscr{H}(\nu_n^{\mathcal{C}} \mid \gamma_n^{t^*}) = 0 \tag{7.8}$$

Of course, in this case, $p^1(t)$ can be computed explicitly; this is most easily done directly from its definition (6.10) using the transfer matrix method (see ref. 13, for example). This yields

$$p^1(t_1, t_2) = \ln\{e^{-t_1} \cosh t_2 + (e^{-2t_1} \sinh^2 t_2 + e^{2t_1})^{1/2}\} \tag{7.9}$$

2b. *The Ising Model* ($d \geqslant 2$): Using (4.31), we saw that the RL-function μ^d exists and is concave; however, in this case it is not strictly concave because there is a first-order phase transition (this was proved by Dobrushin[5] and Griffiths[12] independently) which manifests itself in a ruled patch R on the entropy surface. Because the grand canonical pressure is strictly convex,[6] this patch fits smoothly into the entropy surface. In the case $d = 2$, the boundary of the projection of the ruled patch on the X plane can be computed using Onsager's formula for the spontaneous magnetization, a proof of which was published by Yang.[32] See Fig. 2.

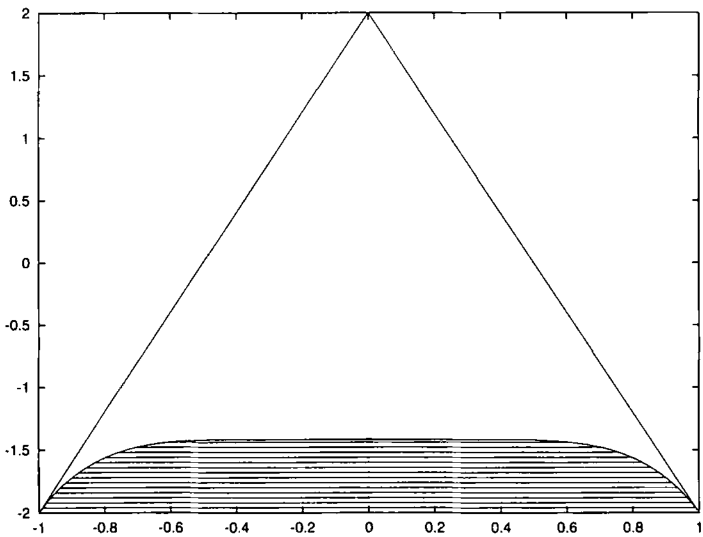

Fig. 2. The projection R of the ruled patch in the entropy surface onto the X plane in the Ising model ($d = 2$).

Choose $C=(c_1', c_1'')\times(c_2', c_2'')$ and suppose that the entropy μ^d is finite at some point of C; the set $X_{\bar{C}}$ is the line-segment $\{c_1''\}\times[c_2', c_2'']\cap R$ provided this is non-empty, and reduces to a single point $\{x^*\}$ otherwise. The set X^t is the single point $\{x_t\}$, where

$$x_t := \operatorname{grad} p(t) \tag{7.10}$$

for values of t for which p is differentiable, and is the line-segment $\{\partial p(t)/\partial t_1\}\times[-m^*(t), m^*(t)]$, where $m^*(t)$ is the spontaneous magnetization, otherwise. We can always choose t^* such that $X_{\bar{C}}\subset X^{t^*}$ and

$$\lim_{n\to\infty}\frac{1}{V_n}\mathscr{H}(\nu_n^C\,|\,\gamma_n^{t^*})=0 \tag{7.11}$$

This follows because, as a consequence of the concavity of μ, we have $\mu(x)=-p^*(x)$ as well as $p(t)=(-\mu)^*(t)$.

3. *The Curie–Weiss Model:* By (6.8), the RL-function μ^{CW} exists; in the case $a\leqslant 1/2$, μ^{CW} is strictly concave and the analysis proceeds as in the case of the paramagnet, establishing that t^* can be chosen so that

$$\lim_{n\to\infty}\frac{1}{V_n}\mathscr{H}(\nu_n^C\,|\,\gamma_n^{t^*})=0 \tag{7.12}$$

However, when $a>1/2$, the RL-function is no longer concave and attains its maximum at two distinct points, x_a and $-x_a$, where x_a is the positive root of

$$\tanh 2ax = x \tag{7.13}$$

Choosing C to be an open interval, we see that the set $X_{\bar{C}}$ consists of one or two points contained in the closed interval $\bar{C}$. The set X^t is, for $t>0$, a single point in the interval $(x_a, 1)$; for $t<0$, a single point in the interval $(-1, -x_a)$; for $t=0$, the pair of points $\{-x_0, x_0\}$. In any case, X^t is a subset of $(-1, -x_0]\cup[x_0, 1)$ and $X_{\bar{C}}$ is a subset of $\bar{C}$, so that, if $\bar{C}\subset(-x_0, x_0)$, there is no value of t for which the condition $X_{\bar{C}}\subset X^t$ is satisfied, since the concentration sets X^t and $X_{\bar{C}}$ are disjoint for all t; one can show that no limit point of a sequence of microcanonical measures is a limit point of a sequence of grand canonical measures.

8. CONCLUSION

We have shown how the criterion of Theorem 6.2 for the vanishing of the specific information gain is satisfied in the model of a paramagnet and in the Ising model with $d=1$ and $d \geqslant 2$. We have demonstrated in the case of the Curie–Weiss model that nonconcavity of the RL-function can prevent its being satisfied. In general, for a lattice gas with translation-invariant summable potentials, the RL-function μ exists and is concave and we refer to it as the *entropy function*; the following result holds: *Let C be an open convex neighborhood of a point at which μ is finite; then there exists t^* such that $X_{\bar{C}} \subset X^{t^*}$ so that*

$$\lim_{n \to \infty} \frac{1}{V_n} \mathscr{H}(\nu_n^C \mid \gamma_n^{t^*}) = 0 \tag{8.1}$$

Using the methods of refs. 30 and 26, we can use this result to prove that *the entropy μ can be used to find a value t^* (generalized chemical potential) such that any weak limit of the sequence $\{\bar{\nu}_n^C\}_{n \geqslant 1}$ is a Gibbs state with respect to the specification determined by $\{\gamma_n^{t^*}\}_{n \geqslant 1}$.*

The existence of a t^* for which $X_{\bar{C}} \subset X^{t^*}$ follows because, as a consequence of the concavity of μ, we have $\mu(x) = -p^*(x)$ as well as $p(t) = (-\mu)^*(t)$; these relations together constitute the equivalence of ensembles at the level of thermodynamic functions. It is in this sense that *in the classical lattice gas, equivalence of ensembles holds at the level of states whenever it holds at the level of thermodynamic functions.* Detailed proofs of these statements will be given in ref. 20.

ACKNOWLEDGMENT

We thank Raymond Russell for writing the code which produced the figures.

REFERENCES

1. M. Aizenmann, S. Goldstein, and J. L. Lebowitz, Conditional equilibrium and the equivalence of microcanonical and grandcanonical ensembles in the thermodynamic limit, *Commun. Math. Phys.* **62**:279–302 (1978).
2. R. R. Bahadur and S. L. Zabel, Large deviations of the sample mean in general vector spaces, *Ann. Prob.* 7:587 (1979).
3. I. Csiszár, Sanov property, generalized I-projection and a conditional limit theorem, *Ann. Prob.* **12**:768–793 (1984).
4. J.-P. Deuschel, D. W. Stroock, and H. Zessin, Microcanonical distribution for lattice gases, *Commun. Math. Phys.* **139**:83 (1991).
5. R. L. Dobrushin, Existence of a phase-transition in two- and three-dimensional Ising models, *Theory Prob. Appl.* **10**:193–213 (1965).
6. R. L. Dobrushin and R. A. Minlos, Existence and continuity of pressure in classical statistical physics, *Theory Prob. Appl.* **12**:535–569 (1967).
7. R. L. Dobrushin and B. Tirozzi, The central limit theorem and the problem of equivalence of ensembles, *Commun. Math. Phys.* **54**:173 (1973).

8. H.-O. Georgii, Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$, *Ann. Prob.*, to appear (1993).
9. H.-O. Georgii, *Canonical Gibbs Measures* (Springer, Berlin, 1979).
10. H.-O. Georgii, Canonical Gibbs states, their relation to Gibbs states, and applications to two-valued Markov chains, *Z. Wahrsch. Verw. Geb.* **32**:277–300 (1975).
11. J. W. Gibbs, *Elementary Principles of Statistical Mechanics* (Yale University Press, New Haven, Connecticut 1902).
12. R. B. Griffiths, Peierls' proof of spontaneous magnetization in a two-dimensional Ising ferromagnet, *Phys. Rev.* **136A**:437–439 (1964).
13. K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
14. E. Ising, Beitrag zur Theorie des Ferromagnetismus, *Z. Phys.* **31**:253–258 (1925).
15. J. H. B. Kemperman, On the optimal rate of transmitting information, in *Probability and Information Theory*, M. Behara, K. Krickeberg, and J. Wolfowitz, eds. (Springer, Berlin, 1969).
16. A. Ya. Khinchin, *Matematicheskie Osnovaniya Statisticheskoi Mekhaniki* (Gostekhizdat, Moscow-Leningrad, 1943) [English translation: *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949)].
17. O. E. Lanford, Entropy and equilibrium states in classical mechanics, in *Statistical Mechanics and Mathematical Problems*, A. Lenard, ed. (Springer, Berlin, 1973).
18. J. T. Lewis and C.-E. Pfister, Thermodynamic probability theory: Some aspects of large deviations, preprint DIAS-93-33 (1993).
19. J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, Large deviations and the thermodynamic formalism: A new proof of the equivalence of ensembles, preprint DIAS-93-24 (1993).
20. J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, Large deviations and the equivalence of ensembles for the classical lattice gas, in preparation.
21. J. T. Lewis and W. G. Sullivan, *Entropy: An Introduction to Statistical Mechanics and the Thermodynamic Formalism* (Wiley, London, in preparation).
22. A. Martin-Löf, *Statistical Mechanics and the Foundations of Thermodynamics* (Springer, Berlin, 1979).
23. A. Martin-Löf, The equivalence of ensembles and Gibbs' phase rule for classical lattice systems, *J. Stat. Phys.* **20**:557–569 (1979).
24. P. Mazur and J. Van der Linden, Asymptotic form of the structure function for real systems, *J. Math. Phys.* **4**:271–277 (1964).
25. M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, (Holden Day, San Francisco, 1964).
26. C. J. Preston, *Random Fields* (Springer, Berlin, 1976).
27. C. J. Preston, Canonical and microcanonical Gibbs states, *Z. Wahrsch. Verw. Geb.* **46**:125–158 (1979).
28. R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).
29. D. Ruelle, Correlation functionals, *J. Math. Phys.* **6**:201 (1965).
30. W. G. Sullivan, Potentials for almost Markovian random fields, *Commun. Math. Phys.* **33**:61 (1973).
31. S. R. S. Varadhan, Asymptotic probabilities and differential equations, *Commun. Pure Appl. Math.* **19**:261 (1966).
32. C. N. Yang, The spontaneous magnetization of a two-dimensional Ising model, *Phys. Rev.* **85**:809–816 (1952).