2000

Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles

RS Ellis
University of Massachusetts - Amherst, rsellis@math.umass.edu

K Haven

B Turkington
University of Massachusetts - Amherst, turk@math.umass.edu

Large Deviation Principles and Complete Equivalence and Nonequivalence Results for Pure and Mixed Ensembles

Richard S. Ellis[*], Kyle Haven[†], and Bruce Turkington[‡]
rsellis@math.umass.edu, haven@math.umass.edu, turk@math.umass.edu
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003

February 8, 2008

Abstract

We consider a general class of statistical mechanical models of coherent structures in turbulence, which includes models of two-dimensional fluid motion, quasi-geostrophic flows, and dispersive waves. First, large deviation principles are proved for the canonical ensemble and the microcanonical ensemble. For each ensemble the set of equilibrium macrostates is defined as the set on which the corresponding rate function attains its minimum of 0. We then present complete equivalence and nonequivalence results at the level of equilibrium macrostates for the two ensembles.

Microcanonical equilibrium macrostates are characterized as the solutions of a certain constrained minimization problem, while canonical equilibrium macrostates are characterized as the solutions of an unconstrained minimization problem in which the constraint in the first problem is replaced by a Lagrange multiplier. The analysis of equivalence and nonequivalence of ensembles reduces to the following question in global optimization. What are the relationships between the set of solutions of the constrained minimization problem that characterizes microcanonical equilibrium macrostates and the set of solutions of the unconstrained minimization problem that characterizes canonical equilibrium macrostates?

In general terms, our main result is that a necessary and sufficient condition for equivalence of ensembles to hold at the level of equilibrium macrostates is that it holds at the level of thermodynamic functions, which is the case if and only if the

[*]This research was supported by a grant from the Department of Energy (DE-FG02-99ER25376) and by a grant from the National Science Foundation (NSF-DMS-9700852).

[†]This research was supported by a grant from the Department of Energy (DE-FG02-99ER25376).

[‡]This research was supported by a grant from the Department of Energy (DE-FG02-99ER25376) and by a grant from the National Science Foundation (NSF-DMS-9971204).

microcanonical entropy is concave. The necessity of this condition is new and has the following striking formulation. If the microcanonical entropy is not concave at some value of its argument, then the ensembles are nonequivalent in the sense that the corresponding set of microcanonical equilibrium macrostates is disjoint from any set of canonical equilibrium macrostates. We point out a number of models of physical interest in which nonconcave microcanonical entropies arise.

We also introduce a new class of ensembles called mixed ensembles, obtained by treating a subset of the dynamical invariants canonically and the complementary set microcanonically. Such ensembles arise naturally in applications where there are several independent dynamical invariants, including models of dispersive waves for the nonlinear Schrödinger equation. Complete equivalence and nonequivalence results are presented at the level of equilibrium macrostates for the pure canonical, the pure microcanonical, and the mixed ensembles.

American Mathematical Society 1991 subject classifications. Primary 60F10, Secondary 82B99

Key words and phrases: Large deviation principle, equilibrium macrostates, equivalence of ensembles, microcanonical entropy

Contents

1 Introduction

1.1 Overview

A wide variety of complex physical systems described by nonlinear partial differential equations exhibit asymptotic phenomena that are much too complicated to study by purely analytic methods. In order to gain a fuller understanding of such phenomena, analytic methods are supplemented by numerical simulations or the systems are modeled via the formalism of statistical mechanics, which often yields uncannily accurate predictions concerning the system's asymptotic behavior.

An important class of complex physical systems for which the formalism of statistical mechanics provides accurate predictions arises in the study of turbulence; e.g., two-dimensional fluid motions, quasi-geostrophic flows, two-dimensional magnetofluids, plasmas, and dispersive waves. In each case important features of the asymptotic behavior of the underlying nonlinear partial differential equation—the two-dimensional Euler equations, the quasi-geostrophic potential vorticity equation, the magnetohydrodynamic equations, the Vlasov-Poisson equation, and the nonlinear Schrödinger equation—can be effectively captured in a statistical mechanical model. A distinguishing feature of such systems is that a free evolution from a generic initial condition exhibits a separation-of-scales behavior: coherent structures are formed on large scales—e.g., vortices and shears in the case of fluid motion or solitons in the case of dispersive waves—while random fluctuations are generated on small scales. A major goal of any description of the system, whether analytic, numeric, or statistical, is to predict the formation, interaction, and persistence of such coherent structures.

The purpose of the present paper is to provide the theoretical basis for statistical mechanical studies of specific models of turbulence that are analyzed elsewhere. These include two-dimensional fluids [6], quasi-geostrophic flows [16], and dispersive waves [17]. In each case the model is defined on a fixed flow domain in terms of a sequence of finite-dimensional systems indexed by $n \in I\!N$. Coherent structures are studied in the continuum limit, obtained by sending $n \to \infty$. They are characterized by variational principles, the solutions of which define equilibrium macrostates. In contrast to the detailed description required by the associated nonlinear partial differential equation and by the finite-dimensional systems that discretize them, these equilibrium macrostates provide a vastly contracted description. The variational principles are derived and analyzed via the theory of large deviations and duality theory for concave functions.

In these models the sequence of finite-dimensional systems is defined on a fixed domain in terms of a long-range interaction with a local mean-field scaling. In order to obtain a nontrivial limit, one must scale the inverse temperature by a parameter tending to infinity. By altering the scaling and making other superficial changes, our results can also be applied to classical lattice models such as the Ising model of a ferromagnet. Such models are typically defined in terms of the thermodynamic limit of a sequence of finite-dimensional systems having a finite-range or summable interaction. In such applications a basic stochastic process that arises in the large deviation analysis is the empirical field, which has been studied by a number of authors including [12, 20, 21, 40]. Other papers that investigate the equivalence of ensembles in the traditional thermodynamic or bulk limit include [1] and [47].

There is a large literature on the equivalence of ensembles for classical lattice systems and related models. It is reviewed in part in the introduction to [33], to which the reader is referred for references. In particular, a number of papers including [12, 21, 32, 46] investigate the equivalence of ensembles using the theory of large deviations. Of these papers, [32] considers the problem in the greatest generality, obtaining a criterion for the equivalence of ensembles in terms of the vanishing of the specific information gain of a sequence of conditioned measures with respect to a sequence of tilted measures. However, despite the mathematical sophistication of these and other studies, none of them explicitly addresses the general issue of the nonequivalence of ensembles, which seems to be the typical behavior for the models of turbulence that the present paper analyzes. In [32, §7.3] and [33, §7] there is a discussion of the nonequivalence of ensembles for the simplest mean-field model in statistical mechanics; namely, the Curie-Weiss model of a ferromagnet. For a general class of local mean-field models of turbulence, the present paper addresses this and related issues.

In much of the classical literature on statistical mechanical approaches to two-dimensional turbulence, it is tacitly assumed that the microcanonical and canonical ensembles give equivalent results [30, 39]. Recently, however, in the context of the point vortex and related models, this tacit assumption has been directly addressed. Questions concerning the equivalence and nonequivalence of ensembles for these models have been investigated by a number of authors, including [9, 19, 26, 28]. The present paper, inspired in part by [19], is the first to present complete and definitive results for a general class of models, with a particular emphasis upon the nonequivalence of ensembles.

An unexpected connection of our work in this paper is to dynamic stability analysis. To date, all studies of the nonlinear stability of two-dimensional flows have been carried out using the Lyapunov functionals introduced by Arnold [2, 3, 35]. When these deterministic results are reformulated in the setting of statistical mechanical models, they can be expressed in terms of the second-order conditions satisfied by canonical equilibrium macrostates. In the cases when the microcanonical entropy is not concave and thus the ensembles are nonequivalent, the Arnold sufficient conditions for nonlinear stability are not satisfied by the microcanonical equilibrium macrostates. Nevertheless, the second-order conditions satisfied by these macrostates allow us to refine the classical Arnold theorems by proving the nonlinear stability of a new class of two-dimensional flows. In [16] these ideas are developed for the quasi-geostrophic potential vorticity equation, which describes the dynamics of rotating, shallow water systems in nearly geostrophic balance. The work in that paper has possible applications to the stability of planetary flows; specifically, to the stability of zonal shear flows and embedded vortices in Jovian-type atmospheres.

In the next two subsections we present an overview of the main results in this paper, stripped of all technicalities. This is done in the context of a well-known statistical mechanical model of the two-dimensional Euler equations known as the Miller-Robert model. Results formulated in great generality to apply to this and other models of turbulence are given in Sections 2-5 of this paper. We start by presenting large deviation principles with respect to the canonical ensemble and the microcanonical ensemble. For each ensemble we then define the set of equilibrium macrostates as the set on which the associated rate function attains its minimum of 0. A fundamental question arises. Are the two ensembles equivalent at the level of equilibrium macrostates? That is, does each equilibrium macrostate with respect to one ensemble correspond to an equilibrium macrostate with respect to the other ensemble? In Section 4, definitive and sharp results on the equivalence and nonequivalence of the ensembles are presented.

In general terms, our main result is that a necessary and sufficient condition for the equivalence of ensembles to hold at the level of equilibrium macrostates is that it holds at the level of thermodynamic functions. In proving this, we go beyond the important work in [32], which proves that for a general class of models including the classical lattice gas thermodynamic equivalence of ensembles is a sufficient condition for macrostate equivalence of ensembles. Our proof that thermodynamic equivalence is also a necessary condition for macrostate equivalence is perhaps the most striking discovery of our work. Specifically, we show that whenever a quantity known as the microcanonical entropy is not concave, the ensembles are nonequivalent in the sense that the set of microcanonical equilibrium macrostates is richer than the set of canonical equilibrium macrostates. In fact, the latter set contains none of the microcanonical equilibrium macrostates corresponding to nonconcave portions of the entropy [see Thm. 4.5(b)]. Useful, but less concrete, connections between the nonconcavity of the microcanonical entropy and nonequivalence of ensembles can also be deduced from the abstract results in [32] [see their §5 and §6]. On the other hand, our results are formulated in order to apply directly to statistical mechanical models of turbulence for which nonconcave microcanonical entropies frequently and naturally arise, particularly in physically interesting regions corresponding to a range of negative temperatures. Several such examples are mentioned in Section 1.4.

Besides the results on equivalence and nonequivalence of ensembles, we also prove that for the Miller-Robert model and other models microcanonical equilibrium macrostates have an equivalent characterization in terms of constrained maximum entropy principles (see Remark 3.4). Our approach to this question seems simpler and more intuitive than the approach taken in [37, 42, 43]. The derivation of constrained maximum entropy principles based on the microcanonical ensemble brings to fruition the work begun in [6], where unconstrained maximum entropy principles based on the canonical ensemble are derived. Our proof that microcanonical equilibrium macrostates are characterized as solutions of constrained maximum entropy principles is an important contribution because such principles are the basis for numerical computations of equilibrium macrostates and coherent structures for the Miller-Robert model and other models [14, 51, 52].

In systems having multiple conserved quantities, one also has the option of studying mixed ensembles. These are defined by treating a subset of the conserved quantities canonically and the complementary subset of conserved quantities microcanonically. In Section 5 we derive large deviation principles with respect to such ensembles and give complete results on their equivalence and nonequivalence, at the level of equilibrium macrostates, with the microcanonical ensemble and the canonical ensemble. Although mixed ensembles arise naturally in a number of applications, they have not been studied in a general setting in the statistical mechanical literature.

An important application of mixed ensembles is to the study of dispersive waves and soliton turbulence for the nonlinear Schrödinger equation [17]. This equation has two conserved quantities, the Hamiltonian and the particle number. In the associated statistical mechanical model, the canonical ensemble cannot be defined because the partition function does not converge. Instead, one must consider either a microcanonical ensemble or a mixed ensemble in which the Hamiltonian is treated canonically and the particle number microcanonically. By applying to the mixed ensemble a large deviation result for Gaussian processes derived in [18], in [17] we are able to justify rigorously a mean-field theoretic approach to soliton turbulence presented in [24]. The agreement between the predictions of the statistical mechanical model and long-time simulations of the microscopic dynamics is excellent [23].

1.2 Ensembles and Large Deviation Principles

The Euler equations describe the time evolution of the velocity field of an inviscid, incompressible fluid in a spatial domain, which for simplicity we take to be the unit torus T^2 with periodic boundary conditions. At time $t > 0$ the velocity field at a position $x = (x_1, x_2) \in T^2$ is denoted $(v_1(x,t), v_2(x,t))$. The Euler equations can be cast in the form of an infinite-dimensional Hamiltonian system having a family of other conserved quantities called generalized enstophies. A central goal of theoretical, numerical, and statistical studies is to relate the asymptotic behavior of the vorticity $\omega(x,t) \doteq v_{2,x_1}(x,t) - v_{1,x_2}(x,t)$ to the formation, interaction, and persistence of coherent structures of the fluid motion.

A model that can be used to carry this out was proposed independently by Miller et. al. [38, 39] and Robert et. al. [43, 44] and is known as the Miller-Robert model. In order to define it, one first discretizes the continuum dynamics described by the Euler equations, and then in terms of the discretized dynamics one defines a sequence of statistical equilibrium models on suitable finite lattices $\mathcal{L}_n$ of T^2. Details are given in part (b) of Example 2.3. These lattice models describe the joint probability distributions of certain vorticity random variables $\zeta(s)$ defined for each site $s \in \mathcal{L}_n$. We denote by ζ the configuration or microstate $\{\zeta(s), s \in \mathcal{L}_n\}$; by a_n the number of sites in $\mathcal{L}_n$; by $\mathcal{Y}$ the common range of $\zeta(s)$; by $H_n(\zeta)$ the Hamiltonian for ζ, which is a certain quadratic function of the $\zeta(s)$ that approximates the continuum Hamiltonian; by $A_n(\zeta)$ the generalized enstrophy of ζ, which approximates the continuum generalized enstrophy; and by P_n the prior distribution of ζ, which is a certain product measure on the configuration space $\mathcal{Y}^{a_n}$. In order to simplify the present description, we absorb A_n in P_n; in [16] a physical justification is given, in the context of a related model, for absorbing the generalized enstrophy A_n in the prior distribution P_n. Thus for the purpose of this introduction, the Miller-Robert model is defined in terms of a single conserved quantity, the Hamiltonian. As in many other models of turbulence, the Hamiltonian in the Miller-Robert model has a long-range interaction and incorporates a local mean-field scaling.

For other models of turbulence having the Hamiltonian as the only conserved quantity, much of the following discussion is valid with minimal changes in notation; in particular, the forms of the large deviation principles in the present subsection and the results on equivalence and nonequivalence of ensembles in the next subsection. For models having multiple conserved quantities, the following discussion is easily adapted by replacing certain scalars with vectors. The general class of models considered in this paper is defined in terms of the quantities in Hypotheses 2.1. In order for a large deviation analysis of the model to be feasible, these quantities must satisfy Hypotheses 2.2.

We begin our overview of the main results in this paper by appealing to the formalism of equilibrium statistical mechanics, which provides two joint probability distributions for microstates $\zeta \in \mathcal{Y}^{a_n}$. The physically fundamental distribution known as the microcanonical ensemble models the fact that the Hamiltonian is a constant of the Euler dynamics. Probabilistically, this is expressed by conditioning P_n on the energy shell $\{\zeta \in \mathcal{Y}^{a_n} : H_n(\zeta) = u\}$, where $u \in I\!\!R$ is determined by the initial conditions. However, in order to avoid problems with the existence of regular conditional probability distributions, we shall condition P_n on the thickened energy shell $\{H_n(\zeta) \in [u-r, u+r]\}$, where $r > 0$. Thus, the microcanonical ensemble is the measure defined for Borel subsets B of $\mathcal{Y}^{a_n}$ by

$$P_n^{u,r}\{B\} = P_n\{B \mid H_n \in [u-r, u+r]\} = \frac{P_n\{B \cap \{H_n \in [u-r, u+r]\}\}}{P_n\{H_n \in [u-r, u+r]\}};$$

this is well defined provided the denominator in the last expression is positive. The letter u is used in the definition of the microcanonical ensemble rather than the more usual letter E because this is a special case of a general theory that applies to models having multiple conserved quantities; for such models $u \in I\!\!R$ is replaced by a vector u representing a fixed value of the vector of conserved quantities.

A mathematically more tractable joint probability distribution is the canonical ensemble, defined for Borel subsets B of $\mathcal{Y}^{a_n}$ by

$$P_{n,\beta}\{B\} \doteq \frac{1}{Z(n,\beta)} \cdot \int_B \exp[-\beta H_n] \, dP_n.$$

Here β is a real number denoting the inverse temperature and $Z(n,\beta)$ is the partition function $\int_{\mathcal{Y}^{a_n}} \exp[-\beta H_n] \, dP_n$. This is a normalization constant that makes $P_{n,\beta}$ a probability measure.

The main mathematical tool that we shall use to predict the formation of coherent structures is the theory of large deviations. In the case of the Miller-Robert model, a crucial innovation implemented in [6] for the canonical ensemble is to study the asymptotic behavior of a random probability measure $Y_n(\zeta)$ that is closely related to a certain coarse graining of the random vorticity field (see part (b) of Example 2.3). This coarse graining is defined in terms of the empirical measures of $\zeta(s)$ for s in certain macrocells of the lattice $\mathcal{L}_n$. Y_n takes values in a certain subset $\mathcal{X}$ of the space of probability measures on $T^2 \times \mathcal{Y}$. Elements μ of $\mathcal{X}$ are called macrostates. While Y_n is basic to analyzing the asymptotic behavior of the model, its definition is far from obvious. For that reason we call Y_n a hidden process and $\mathcal{X}$ a hidden space for the Miller-Robert model.

The hidden process Y_n has two properties that make a large deviation analysis of the Miller-Robert model possible. For details, the reader is referred to [6]. First, an application of Sanov's Theorem shows that with respect to the a priori distribution P_n, Y_n satisfies the large deviation principle on $\mathcal{X}$ with rate function $I(\mu)$ given by the relative entropy of $\mu \in \mathcal{X}$ with respect to a certain base measure. We record this fact by the formal notation

$$P_n\{Y_n \in B(\mu, \alpha)\} \approx \exp[-a_n I(\mu)] \ \text{ as } n \to \infty, \alpha \to 0. \tag{1.2.1}$$

In this formula $B(\mu, \alpha)$ denotes the open ball with center μ and radius α with respect to an appropriate metric on $\mathcal{X}$. Second, there exists a bounded continuous function $\tilde{H}$ mapping $\mathcal{X}$ into $I\!R$ with the property that uniformly over microstates the Hamiltonian $H_n(\zeta)$ is asymptotic to $\tilde{H}(Y_n(\zeta))$ as $n \to \infty$; in symbols,

$$\lim_{n\to\infty} \sup_{\zeta \in \mathcal{Y}^{a_n}} |H_n(\zeta) - \tilde{H}(Y_n(\zeta))| = 0. \tag{1.2.2}$$

$\tilde{H}$ is called the Hamiltonian representation function.

Using (1.2.2), one derives from the large deviation principle for the P_n-distributions of Y_n the asymptotic behavior of Y_n with respect to the two ensembles $P_n^{u,r}$ and $P_{n,a_n\beta}$. For appropriate values of u and β these are expressed by the formal notation

$$P_n^{u,r}\{Y_n \in B(\mu, \alpha)\} \approx \exp[-a_n I^u(\mu)] \ \text{ as } n \to \infty, r \to 0, \alpha \to 0 \tag{1.2.3}$$

and

$$P_{n,a_n\beta}\{Y_n \in B(\mu, \alpha)\} \approx \exp[-a_n I_\beta(\mu)] \ \text{ as } n \to \infty, \alpha \to 0. \tag{1.2.4}$$

In these formulas I^u and I_β are rate functions that map $\mathcal{X}$ into $[0, \infty]$ and are defined in terms of the relative entropy I appearing in (1.2.1). Because the Miller-Robert model is defined in terms of a long-range interaction having a local mean-field scaling, in order to obtain a nontrivial asymptotic theory β must be scaled by a_n in the definition of the canonical ensemble $P_{n,\beta}$ [6, §3]. For the general formulation of (1.2.3) and (1.2.4) as large deviation principles for a general class of models, the reader is referred to Theorem 3.2 and Theorem 2.4, respectively.

It is not difficult to motivate the forms of I^u and I_β. In order to do so, we introduce two basic thermodynamic functions, one associated with each ensemble. Since the ground-breaking work of Lanford on equilibrium macrostates in classical statistical mechanics [31], it has been recognized that the basic thermodynamic function associated with the microcanonical ensemble is the microcanonical entropy s. In terms of the distribution $P_n\{H_n \in \cdot\}$, this quantity measures the multiplicity of microstates $\zeta \in \mathcal{Y}^{a_n}$ consistent with a given energy value u. It is defined by

$$s(u) \doteq \lim_{r\to 0} \lim_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in [u - r, u + r]\}. \tag{1.2.5}$$

For appropriate values of u, the limit exists and is given by (3.2), which is a variational formula over macrostates μ. For $\beta \in I\!R$ the basic thermodynamic function associated with the canonical ensemble is the canonical free energy

$$\varphi(\beta) \doteq -\lim_{n\to\infty} \frac{1}{a_n} \log Z(n, a_n\beta). \tag{1.2.6}$$

The limit exists and is given by (2.6), which is also a variational formula over macrostates.

We first motivate the form of I_β. If $Y_n \in B(\mu,\alpha)$, then for all sufficiently small α and all sufficiently large n (1.2.2) implies that

$$H_n(\zeta) \approx \tilde{H}(Y_n(\zeta)) \approx \tilde{H}(\mu).$$

Hence for all sufficiently small α and all sufficiently large n, the asymptotic formula (1.2.1) and the definition of φ yield

$$\begin{aligned} P_{n,a_n\beta}\{Y_n \in B(\mu,\alpha)\} &\doteq \frac{1}{Z(n,\beta)} \int_{\{Y_n \in B(\mu,\alpha)\}} \exp[-a_n\beta H_n]\, dP_n \\ &\approx \frac{1}{Z(n,\beta)} \exp[-a_n\beta\tilde{H}(\mu)]\, P_n\{Y_n \in B(\mu,\alpha)\} \\ &\approx \exp[-a_n(I(\mu)+\beta\tilde{H}(\mu)-\varphi(\beta))]. \end{aligned}$$

Comparing this with the desired asymptotic form (1.2.4) motivates the formula

$$I_\beta(\mu) = I(\mu) + \beta\tilde{H}(\mu) - \varphi(\beta). \tag{1.2.7}$$

The actual proof of the large deviation principle for the $P_{n,a_n\beta}$-distributions of Y_n with this rate function follows the sketch presented here and is not difficult. Related large deviation principles have been obtained by numerous authors.

We now motivate the form of I^u. Suppose that $\tilde{H}(\mu) = u$. Then for all sufficiently large n depending on r the set of ζ for which both $Y_n(\zeta) \in B(\mu,\alpha)$ and $H_n(\zeta) \in [u-r, u+r]$ is approximately equal to the set of ζ for which both $Y_n(\zeta) \in B(\mu,\alpha)$ and $\tilde{H}(Y_n(\zeta)) \in [u-r, u+r]$. Since $\tilde{H}$ is continuous and $\tilde{H}(\mu) = u$, for all sufficiently small α compared to r this set reduces to $\{\zeta : Y_n(\zeta) \in B(\mu,\alpha)\}$. Hence for all sufficiently small r, all sufficiently large n depending on r, and all sufficiently small α compared to r, (1.2.1) and the definition (1.2.5) of s yield

$$\begin{aligned} P_n^{u,r}\{Y_n \in B(\mu,\alpha)\} &\doteq \frac{P_n\{\{Y_n \in B(\mu,\alpha)\} \cap \{H_n \in [u-r,u+r]\}\}}{P_n\{H_n \in [u-r,u+r]\}} \\ &\approx \frac{P_n\{Y_n \in B(\mu,\alpha)\}}{P_n\{H_n \in [u-r,u+r]\}} \\ &\approx \exp[-a_n(I(\mu)+s(u))]. \end{aligned}$$

On the other hand, if $\tilde{H}(\mu) \neq u$, then a similar calculation shows that for all sufficiently small r, all sufficiently small α, and all sufficiently large n $P_n^{u,r}\{Y_n \in B(\mu,\alpha)\} = 0$. Comparing these approximate calculations with the desired asymptotic form (1.2.3) motivates the formula

$$I^u(\mu) \doteq \begin{cases} I(\mu) + s(u) & \text{if } \tilde{H}(\mu) = u, \\ \infty & \text{if } \tilde{H}(\mu) \neq u. \end{cases} \tag{1.2.8}$$

In Section 3 we offer two proofs of the large deviation principle for the $P_n^{u,r}$-distributions of Y_n. Both are straightforward; the first follows fairly closely the heuristic sketch just given. Forms of this large deviation principle are given, for example, in [12, 32, 33].

The asymptotic formulas (1.2.3) and (1.2.4) give rise to several interpretations of the rate functions. Through the distributions $P_n^{u,r}\{Y_n \in \cdot\}$ and $P_{n,a_n\beta}\{Y_n \in \cdot\}$, I^u and I_β measure the multiplicity of microstates $\zeta \in \mathcal{Y}^{a_n}$ consistent with a given macrostate μ. Because of these asymptotic formulas, it also makes sense to say that for $i = I^u$ or $i = I_\beta$ a macrostate $\mu_1 \in \mathcal{X}$ is more predictable than a macrostate $\mu_2 \in \mathcal{X}$ if $i(\mu_1) < i(\mu_2)$. Since i is nonnegative, the most predictable or most probable macrostates μ solve $i(\mu) = 0$. It is natural to call such μ equilibrium macrostates. Specifically, $\mu \in \mathcal{X}$ satisfying $I^u(\mu) = 0$ is called a microcanonical equilibrium macrostate; $\mathcal{E}^u$ denotes the set of all such macrostates. Analogously, a measure $\mu \in \mathcal{X}$ satisfying $I_\beta(\mu) = 0$ is called a canonical equilibrium macrostate; $\mathcal{E}_\beta$ denotes the set of all such macrostates. In terms of equilibrium macrostates μ, one can analyze the formation of coherent structures by defining the mean vorticity as an appropriate average of μ and comparing it, say by simulation, with the long-time behavior of the vorticity $\omega(x,t) \doteq v_{2,x_1}(x,t) - v_{1,x_2}(x,t)$ as given by the Euler equations [39, 44, 51, 52].

1.3 Equivalence and Nonequivalence of Ensembles

The microcanonical ensemble is physically fundamental, and the canonical ensemble can be heuristically derived from it by considering a small subsystem of a large reservoir [4]. Aside from physical considerations concerning which ensemble is more appropriate in the construction of a statistical model, the more mathematically tractable canonical ensemble is often introduced as an approximation to the microcanonical ensemble, which is somewhat difficult to analyze. However, in order to justify this use of the canonical ensemble, one must address a basic issue. At the level of equilibrium macrostates, do the two ensembles give equivalent results? This involves answering the following two questions.

1. For every β and every μ in the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates, does there exist a value of u such that μ lies in the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates?

2. Conversely, for every u and every $\mu \in \mathcal{E}^u$ does there exist a value of β such that $\mu \in \mathcal{E}_\beta$?

Whether or not the answers are yes, a more refined issue is to determine the precise relationships between $\mathcal{E}^u$ and $\mathcal{E}_\beta$. For example, if the answers are both yes, then given β in question 1 (resp., u in question 2), how does one determine the corresponding value of u (resp., β)? It is with these issues, appropriately formulated in terms of a general class of models having multiple conserved quantities, that Sections 4 and 5 of the present paper is occupied. In those sections definitive and sharp results on the equivalence and nonequivalence of ensembles are derived.

As we will see, in general question 1 in the preceding paragraph has the answer yes; namely, every $\mu \in \mathcal{E}_\beta$ lies in $\mathcal{E}^u$ for some value of u. As we illustrate by a number of examples given in Section 1.4, question 2 can have the answer no; namely, it can be the case that the set of microcanonical equilibrium macrostates is richer than the set of canonical equilibrium macrostates. As we show in Theorem 4.4, this behavior has a striking formulation in terms of the microcanonical entropy s, which is defined in (1.2.5). If s is not concave at a given value of u, then the ensembles are nonequivalent in the sense that $\mathcal{E}^u$ is disjoint from the sets $\mathcal{E}_\beta$ for all values of β.

This general result has been anticipated in a number of works, including those discussed in Section 4.2 of [49] and in [27, 29]. These works exhibit nonconcave entropy curves for a number of physical models that include a gravitating system of fermions and a system of circular vortex filaments in an ideal fluid confined to a three-dimensional torus; see Fig. 34 in [49], Fig. 3 in [27], and Fig. 2 in [29]. They also point out that certain equilibrium macrostates corresponding to nonconcave portions of the entropy are only realizable in the continuum limit of the microcanonical ensemble but not of the canonical ensemble. Other examples of nonconcave entropies are given in Section 1.4 of the present paper.

The question as to whether the microcanonical and canonical ensembles give equivalent results at the level of equilibrium macrostates is formulated as a problem in global optimization. Let u and β be given. By definition, a macrostate $\bar{\mu}$ belongs to $\mathcal{E}^u$ if and only if $I^u(\bar{\mu}) = 0$. This is the case if and only if $\bar{\mu}$ solves the following constrained minimization problem:

$$\text{minimize } I(\mu) \text{ over } \mu \in \mathcal{X} \text{ subject to the constraint } \tilde{H}(\mu) = u; \tag{1.3.1}$$

it is worth noting that since the relative entropy $I(\mu)$ equals negative the physical entropy, this display defines a maximum entropy principle with the energy constraint $\tilde{H}(\mu) = u$. By definition, a macrostate $\bar{\mu}$ belongs to $\mathcal{E}_\beta$ if and only if $I_\beta(\bar{\mu}) = 0$. This is the case if and only if $\bar{\mu}$ solves the following unconstrained minimization problem:

$$\text{minimize } (I(\mu) + \beta\tilde{H}(\mu)) \text{ over } \mu \in \mathcal{X}. \tag{1.3.2}$$

In the unconstrained problem β is a Lagrange multiplier dual to the constraint $\tilde{H}(\mu) = u$ in (1.3.1). Under general conditions, solutions of the constrained minimization problem (1.3.1) are extremal points of $(I + \beta\tilde{H})$ on $\mathcal{X}$ [22, 53]. The question as to whether the microcanonical and canonical ensembles give equivalent results is equivalent to answering the following refined question related to this property. What are the relationships between the sets of solutions of the constrained and unconstrained minimization problems (1.3.1) and (1.3.2)?

We now describe our results on the equivalence and nonequivalence of ensembles by relating them to the behavior of the two basic thermodynamic functions, s and φ. The following discussion applies to the Miller-Robert model as well as to a class of other models that have the Hamiltonian as a single conserved quantity. The discussion generalizes to a wide class of other models having multiple conserved quantities. We first motivate a formula relating s and φ. To do this, we use the definition of s, which we summarize by the formula

$$P_n\{H_n \in du\} \approx \exp[a_n s(u)]\, du.$$

We now calculate

$$\varphi(\beta) \quad = \quad -\lim_{n\to\infty} \frac{1}{a_n} \log Z(n, a_n\beta)$$

$$
\begin{aligned}
&= -\lim_{n\to\infty}\frac{1}{a_n}\log\int_{\mathcal{Y}^{a_n}}\exp[-a_n\beta H_n]\,dP_n \\
&= -\lim_{n\to\infty}\frac{1}{a_n}\log\int_{I\!R}\exp[-a_n\beta u]\,P_n\{H_n\in du\} \\
&\approx -\lim_{n\to\infty}\frac{1}{a_n}\log\int_{I\!R}\exp[-a_n(\beta u - s(u))]\,du.
\end{aligned}
$$

According to the heuristic reasoning that underlies Laplace's method, the main contribution to the integral comes from the largest term. This motivates the relationship

$$\varphi(\beta) = \inf_{u\in I\!R}\{\beta u - s(u)\}, \tag{1.3.3}$$

which expresses φ as the Legendre-Fenchel transform s^* of s.

For the Miller-Robert model and other models of turbulence considered in this paper, s is nonpositive and upper semicontinuous on $I\!R$ [Prop. 3.1(a)]. If it is the case that s is concave on $I\!R$, then (1.3.3) can be inverted to give s in terms of φ; namely, for all $u \in I\!R$

$$s(u) = \inf_{\beta\in I\!R}\{\beta u - \varphi(\beta)\}. \tag{1.3.4}$$

Hence, when s is concave on $I\!R$, each basic thermodynamic function can be obtained from the other by a similar formula. It is natural to say that in this case the microcanonical ensemble and the canonical ensemble are thermodynamically equivalent [28, 33]. As we will see in Theorems 4.4 and 4.9, thermodynamic equivalence of ensembles is mirrored by equivalence-of-ensemble relationships at the level of equilibrium macrostates.

By virtue of its definition (1.2.6) or formula (1.3.3), φ is a finite, concave, continuous function on $I\!R$. In the case of classical systems such as considered by Lanford [31], a superadditivity argument based on the fact that the underlying Hamiltonian has finite range shows that the analogue of s is an upper semicontinuous, concave function on $I\!R$. In general, however, because of the local mean-field, long-range nature of the Hamiltonians in the Miller-Robert model and other models of turbulence considered in this paper, the associated microcanonical entropies are typically not concave on subsets of $I\!R$ corresponding to a range of negative temperatures.

In order to see how concavity properties of s determine relationships between the sets of equilibrium macrostates, we define for $u \in I\!R$ the concave function

$$s^{**}(u) \doteq \inf_{\beta\in I\!R}\{\beta u - s^*(\beta)\} = \inf_{\beta\in I\!R}\{\beta u - \varphi(\beta)\}. \tag{1.3.5}$$

Because of (1.3.4), it is obvious that s is concave on $I\!R$ if and only if s and s^{**} coincide. Whenever $s(u) > -\infty$ and $s(u) = s^{**}(u)$, we shall say that s is concave at u.

Now assume that s is not concave on $I\!R$; i.e., there exists $u \in I\!R$ for which $-\infty < s(u) \neq s^{**}(u)$. In this case, one easily shows that s^{**} equals the smallest upper semicontinuous, concave function majorizing s. In particular, when s is not concave on $I\!R$, it cannot be recovered from φ via a Legendre-Fenchel transform.

As we now explain, concavity and nonconcavity properties of the microcanonical entropy s have crucial implications for the equivalence and nonequivalence of ensembles at the level of equilibrium macrostates. In terms of such properties of s, we now give preliminary and incomplete statements of the relationships between the sets $\mathcal{E}^u$ and $\mathcal{E}_\beta$ of equilibrium macrostates for the two ensembles. The reader is referred to Theorems 4.4, 4.6, and 4.8 for precise statements. For easy reference they are summarized in Figure 1 in Section 4.

For a given value of u, there are three possible relationships that can occur between $\mathcal{E}^u$ and $\mathcal{E}_\beta$. If there exists a value of β such that $\mathcal{E}^u = \mathcal{E}_\beta$, then the ensembles are said to be fully equivalent. If instead of equality $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$ for some β, then the ensembles are said to be partially equivalent. It may also happen that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all values of β. If this occurs, then the microcanonical ensemble is said to be nonequivalent to any canonical ensemble or that nonequivalence of ensembles holds. It is convenient to group the first two cases together. If for a given u there exists β such that either $\mathcal{E}^u$ equals $\mathcal{E}_\beta$ or $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$, then the ensembles are said to be equivalent.

The relationships between $\mathcal{E}^u$ and $\mathcal{E}_\beta$ depend on concavity and nonconcavity properties of s, expressed through the equality or nonequality of $s(u)$ and $s^{**}(u)$. These relationships are given next in items 1-3 together with references to where the results are stated precisely. Criteria for equivalence of ensembles related to item 2 have been obtained in various settings by a number of authors, including [12, 19, 32, 33]. However, the results underlying items 1 and 3 are new.

1. **Canonical is always microcanonical.** For every β and every $\mu \in \mathcal{E}_\beta$, there exists u such that $\mu \in \mathcal{E}^u$ [Theorem 4.6].

2. **Equivalence.** If $-\infty < s(u) = s^{**}(u)$—i.e., if s is concave at u—then there exists β such that the ensembles are equivalent [Remark 4.2 and Theorem 4.4(a)].

3. **Nonequivalence.** If $-\infty < s(u) \neq s^{**}(u)$—i.e., if s is not concave at u—then the corresponding microcanonical ensemble is nonequivalent to any canonical ensemble [Remark 4.2 and Theorem 4.4(b)].

Let u be a point in $I\!\!R$ such that $s(u) > -\infty$. According to items 2 and 3, the ensembles are equivalent if and only if s is concave at u. Under another natural hypothesis on u, one shows that s is concave at u if and only if there exists a supporting line to the graph of s at $(u, s(u))$ [Lem. 4.1(a)]; i.e., there exists $\beta \in I\!\!R$ such that

$$s(w) \leq s(u) + \beta(w - u) \ \text{ for all } w \in I\!\!R.$$

In Theorem 4.8 we refine this necessary and sufficient condition for equivalence of ensembles by showing that the ensembles are fully equivalent if and only if there exists a supporting line to the graph of s that touches the graph of s only at $(u, s(u))$; i.e., there exists $\beta \in I\!\!R$ such that

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

A sufficient condition that guarantees this property of s is that $s(u) = s^{**}(u)$ and s^{**} is strictly concave in a neighborhood of u.

The relationships given in items 1-3 refine the relationships between the thermodynamic functions φ and s. In fact, the thermodynamic equivalence of ensembles that holds when $s = s^{**}$ on $I\!R$ is reflected in the equivalence of ensembles for a given value of u when $-\infty < s(u) = s^{**}(u)$ [item 2]. On the other hand, when $-\infty < s(u) \neq s^{**}(u)$ for some value of u, the lack of symmetry between φ and s as expressed by (1.3.3) and (1.3.5) is mirrored by a lack of symmetry between the microcanonical and canonical ensembles at the level of equilibrium macrostates. For each β, every canonical equilibrium macrostate in $\mathcal{E}_\beta$ lies in $\mathcal{E}^u$ for some u [item 1]. However, for any u for which $-\infty < s(u) \neq s^{**}(u)$ the corresponding microcanonical ensemble is nonequivalent to any canonical ensemble [item 3].

We also prove a number of interesting results that follow easily from the main theorems. For example, in Corollary 4.7 we show that if $\mathcal{E}_\beta$ consists of a unique macrostate μ, then $\mathcal{E}^u$ consists of the unique macrostate μ for a corresponding value of u ($u = \tilde{H}(\mu)$). The uniqueness of an equilibrium macrostate corresponds to the absence of a phase transition.

1.4 Examples of Nonconcave Microcanonical Entropies

The most striking of our results on equivalence and nonequivalence of ensembles is given in item 3 near the end of the preceding subsection. If, for a given value of u, $-\infty < s(u) \neq s^{**}(u)$, then $\mathcal{E}^u$ is disjoint from the sets $\mathcal{E}_\beta$ for all values of β. We next point out a number of statistical mechanical models having a nonconcave microcanonical entropy and thus exhibiting, for a range of values of u, the nonequivalence of ensembles that is formulated in item 3.

1. **Point vortex system.** This is the first statistical mechanical model proposed in the literature for studying the two-dimensional Euler equations. It is defined in terms of a singular interaction function, which is a Green's function. The model was introduced by Onsager [41]; was further developed in the 1970's, notably by Joyce and Montgomery [25]; and continues to be the subject of important studies, including [5, 8, 9, 26, 28]. Proposition 6.2 in [9] isolates a class of flow domains for which the microcanonical entropy in the point vortex model is not a concave function of its argument. As pointed out in [28, §6], the Monte Carlo study of a point vortex system in a disk carried out in [48] also displays a nonconcave microcanonical entropy. Strictly speaking, the results on nonequivalence of ensembles given in the present paper apply only to a point vortex model in which the singular interaction function in the classical model has been regularized; see part (a) of Example 2.3. Nevertheless, special arguments can be invoked to extend them to the classical model with singular point vortices.

2. **Two-dimensional turbulence.** A natural generalization, and also regularization, of the point vortex model is the Miller-Robert model. In an unpublished numerical study, Turkington and Liang consider the Miller-Robert model in a disk with constraints on the energy, the total circulation, and the angular momentum (or impulse) and with a prior distribution on the vorticity that corresponds to vortex patch dynamics; this problem is the simplest Miller-Robert analogue of the problem

studied in [48] in the point-vortex formulation. For fixed values of the total circulation and the angular momentum, Turkington and Liang compute microcanonical entropies as a function of energy using the algorithm developed in [51]. They find that the microcanonical entropy-energy curve is concave on a certain interval and nonconcave on a complementary interval. These computations produce equilibrium macrostates that are vortices embedded in circular shear flows.

3. **Quasi-geostrophic turbulence on a β-plane.** The statistical equilibrium models proposed in [50] are implemented in [14] for barotropic, quasi-geostrophic flow in a channel on the β-plane. Various prior distributions on the potential vorticity are considered; these include a saturated model, in which the maximum and minimum of the potential vorticity constrain the microstate, and a dilute model, in which only the mean potential-vorticity magnitude is imposed. Even in the absense of geophysical effects ($\beta = 0$), the dilute model exhibits a nonconcave entropy-energy curve, as displayed in Figure 4 of [14]. The equilibrium macrostates corresponding to values of the energy for which the entropy is nonconcave are shears that transition to monopolar vortices and then to dipolar vortices as the energy increases. When the dilute model is replaced by the corresponding saturated model, in which an upper bound on the microscopic potential vorticity is enforced, the equilibrium macrostates are modified, particularly at high energies. As is shown in Figure 16 of [14], the nonconcavity of the entropy-energy curve persists at low energies; at high enough energies, however, it becomes concave, unlike in the dilute case. At these high energies the equilibrium macrostates are not dipolar vortices, but rather shear flows.

4. **Quasi-geostrophic turbulence over topography.** A more complete study of the concavity of the microcanonical entropy is carried out in [16] for equivalent-barotropic, quasi-geostrophic flow over bottom topography on an f-plane. As in [14] a channel geometry is imposed, but for simplicity only shear flows are considered. Within this symmetry class, the topography is chosen to be sinusoidal, the energy and circulation are used as global invariants, and the prior distribution is taken to be a Gamma distribution with mean 0, variance 1, and nonzero skewness. As a function of the energy and the circulation, the entropy is nonconcave in more than half of its domain. These two-constraint results are described in detail in Section 6 of [16].

5. **Two-layer quasi-geostrophic turbulence.** The one-layer model studied in [14] is extended to a two-layer system in [13], where it is used to describe the physically important phenomenon of open-ocean convection. In Figures 2 and 12 in that paper, the entropy-energy curve is seen to be nonconcave; the microcanonical equilibrium macrostates corresponding to values of the energy in the nonconcave region are asymmetric baroclinic monopoles.

1.5 Contents of This Paper

In Section 2 we introduce the class of statistical mechanical models that will be analyzed in this paper. These models generalize the Miller-Robert model by incorporating a finite sequence of interaction functions $H_{n,i}$ rather than just the Hamiltonian. In order to carry out the large deviation analysis, we assume that there exists a hidden process Y_n that takes values in a complete separable metric space $\mathcal{X}$ and has the following two properties: (a) for each interaction function there exists a representation function $\tilde{H}_i$ such that uniformly over microstates $|H_{n,i} - \tilde{H}_i \circ Y_n| \to 0$ as $n \to \infty$; (b) with respect to the prior measure P_n in the model, Y_n satisfies the large deviation principle on $\mathcal{X}$. In Section 2 we show that with respect to the canonical ensemble Y_n satisfies the large deviation principle, and we derive several properties of the set of canonical equilibrium macrostates.

In Section 3 we consider the microcanonical ensemble, proving a large deviation principle and studying properties of the set of microcanonical equilibrium macrostates. We also point out the constrained maximum entropy principles that characterize microcanonical equilibrium macrostates in certain models including the Miller-Robert model.

Section 4 is devoted to the presentation of our complete results on the equivalence and nonequivalence of the two ensembles. The results are proved in Theorems 4.4, 4.6, and 4.8 and are summarized in Figure 1.

In Section 5.1 we introduce mixed ensembles obtained by treating a subset of the dynamical invariants canonically and the complementary subset of dynamical invariants microcanonically. We then prove the large deviation principle for these ensembles. Section 5.2 presents complete equivalence and nonequivalence results for the pure canonical and mixed ensembles while Section 5.3 does the same for the mixed and the pure microcanonical ensembles. The results in Sections 5.2 and 5.3 follow from those in Section 4 with minimal changes in proof. They are summarized in Figures 2 and 3.

Acknowledgement. We thank Michael Kiessling for a number of useful conversations.

2 Canonical Ensemble: LDP and Equilibrium Macrostates

In this section we present a large deviation principle for the canonical ensemble in a wide range of statistical mechanical models [Thm. 2.4(b)]. In terms of that principle, the set of canonical equilibrium macrostates is defined and some of its properties derived [Thms. 2.4(c)-2.5]. After defining the class of models under consideration, we specify in Example 2.3 a number of specific models to which the theory applies.

The models that we consider are defined in terms of the following quantities.

Hypotheses 2.1.

- A sequence of probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ indexed by $n \in I\!N$; Ω_n are the configuration spaces for the statistical mechanical models.

- A positive integer σ and for each $n \in I\!\!N$ a sequence of interaction functions $\{H_{n,i}, i = 1, \ldots, \sigma\}$, which are bounded measurable functions mapping Ω_n into $I\!\!R$. We define $H_n \doteq (H_{n,1}, \ldots, H_{n,\sigma})$, which maps Ω_n into $I\!\!R^\sigma$.

- A sequence of positive scaling constants $a_n \to \infty$.

Let $\langle \cdot, \cdot \rangle$ denote the Euclidean inner product on $I\!\!R^\sigma$. We define for each $n \in I\!\!N$, $\beta = (\beta_1, \ldots, \beta_\sigma) \in I\!\!R^\sigma$, and set $B \in \mathcal{F}_n$ the partition function

$$Z_n(\beta) \doteq \int_{\Omega_n} \exp\left[- \sum_{i=1}^{\sigma} \beta_i H_{n,i} \right] dP_n = \int_{\Omega_n} \exp[-\langle \beta, H_n \rangle] \, dP_n,$$

which is well defined and finite, and the probability measure

$$P_{n,\beta}\{B\} \doteq \frac{1}{Z_n(\beta)} \int_B \exp[-\langle \beta, H_n \rangle] \, dP_n. \tag{2.1}$$

The measures $P_{n,\beta}$ are Gibbs states that define the canonical ensemble for the given model. For $\beta \in I\!\!R^\sigma$, we also define

$$\varphi(\beta) \doteq - \lim_{n \to \infty} \frac{1}{a_n} \log Z_n(a_n \beta)$$

if the limit exists and is nontrivial. In this formula β is scaled with a_n, as is usual in studying the continuum limit of models of turbulence [6, §3]. We refer to $\varphi(\beta)$ as the canonical free energy. If $\sigma = 1$ and $H_{n,1}$ is the Hamiltonian of the system, then $\beta = \beta_1$ is the inverse temperature.

The first application of the theory of large deviations in this paper is to express $\varphi(\beta)$ as a variational formula. Let $\mathcal{X}$ be a Polish space (a complete separable metric space), Y_n random variables mapping Ω_n into $\mathcal{X}$, Q_n probability measures on $(\Omega_n, \mathcal{F}_n)$, and I a rate function on $\mathcal{X}$. Thus I maps $\mathcal{X}$ into $[0, \infty]$ and for each $M \in [0, \infty)$ the set $\{x \in \mathcal{X} : I(x) \leq M\}$ is compact (compact level sets). For A a subset of $\mathcal{X}$, we define $I(A) \doteq \inf_{x \in A} I(x)$. We say that with respect to Q_n the sequence Y_n satisfies the large deviation principle, or LDP, on $\mathcal{X}$ with scaling constants a_n and rate function I if for any closed subset F of $\mathcal{X}$ the large deviation upper bound

$$\limsup_{n \to \infty} \frac{1}{a_n} \log Q_n\{Y_n \in F\} \leq -I(F) \tag{2.2}$$

is valid and for any open subset G of $\mathcal{X}$ the large deviation lower bound

$$\liminf_{n \to \infty} \frac{1}{a_n} \log Q_n\{Y_n \in F\} \geq -I(G) \tag{2.3}$$

is valid. We say that with respect to Q_n the sequence Y_n satisfies the Laplace principle on $\mathcal{X}$ with scaling constants a_n and rate function I if for all bounded continuous functions f mapping $\mathcal{X}$ into $I\!\!R$

$$\begin{aligned} &\lim_{n \to \infty} \frac{1}{a_n} \log \int_{\Omega_n} \exp[a_n f(Y_n)] \, dQ_n \\ &\quad = \lim_{n \to \infty} \frac{1}{a_n} \log \int_{\mathcal{X}} \exp[a_n f(x)] \, Q_n\{Y_n \in dx\} = \sup_{x \in \mathcal{X}} \{f(x) - I(x)\}. \end{aligned}$$

As pointed out in Theorems 1.2.1 and 1.2.3 in [15], Y_n satisfies the LDP with scaling constants a_n and rate function I if and only if Y_n satisfies the Laplace principle with scaling constants a_n and rate function I. Evaluating the large deviation upper bound (2.2) for $F = \mathcal{X}$ and the large deviation lower bound (2.3) for $G = \mathcal{X}$ yields $I(\mathcal{X}) = 0$, and since I is nonnegative and has compact level sets, the set of $x \in \mathcal{X}$ for which $I(x) = 0$ is nonempty and compact. In the sequel we shall usually omit the phrase "with scaling constants a_n" in the statements of LDP's and Laplace principles.

A large deviation analysis of the general model is possible provided we can find, as specified in Hypotheses 2.2, a hidden space, a hidden process, and a sequence of interaction representation functions, and provided the hidden process satisfies the LDP on the hidden space.

Hypotheses 2.2.

- **Hidden space.** This is a Polish space $\mathcal{X}$.
- **Hidden process.** This is a sequence Y_n, where each Y_n is a random variable mapping Ω_n into $\mathcal{X}$.
- **Interaction representation functions.** This is a sequence $\{\tilde{H}_i, i = 1, \ldots, \sigma\}$ of bounded continuous functions mapping $\mathcal{X}$ into $I\!R$ such that as $n \to \infty$

$$H_{n,i}(\omega) = \tilde{H}_i(Y_n(\omega)) + \mathrm{o}(1) \text{ uniformly for } \omega \in \Omega_n; \tag{2.4}$$

  i.e., $\lim_{n\to\infty} \sup_{\omega\in\Omega_n} \left| H_{n,i}(\omega) - \tilde{H}_i(Y_n(\omega)) \right| = 0$. We define $\tilde{H} \doteq (\tilde{H}_1, \ldots, \tilde{H}_\sigma)$, which maps $\mathcal{X}$ into $I\!R^\sigma$.
- **LDP for the hidden process.** There exists a rate function I mapping $\mathcal{X}$ into $[0, \infty]$ such that with respect to P_n the sequence Y_n satisfies the LDP on $\mathcal{X}$, or equivalently the Laplace principle on $\mathcal{X}$, with rate function I.

In this context we use the term "hidden" because in many cases the choices of the space $\mathcal{X}$ and the process Y_n are far from obvious. ■

We next present several models of turbulence to which the results of this paper can be applied.

Example 2.3. (a) **Regularized Point Vortex Model.** This model, analyzed in [19], is an approximation to the point vortex model, which we first define. Let Λ be a smooth, bounded, connected, open subset of $I\!R^2$; $g(x, x')$ the Green's function for $-\triangle$ on Λ with Dirichlet boundary conditions; h the continuous function mapping Λ into $I\!R$ defined by $h(x) \doteq \frac{1}{2}\tilde{g}(x, x)$, where $\tilde{g}(x, x')$ is the regular part of the Green's function $g(x, x')$; and θ normalized Lebesgue measure on Λ satisfying $\theta(\Lambda) = 1$. For $n \in I\!N$ the point vortex model is defined on the configuration spaces $\Omega_n \doteq \Lambda^n$ with the Borel σ-field. P_n equals the product measure on Ω_n with identical one-dimensional marginals θ, and $a_n \doteq n$.

Configurations $\zeta \in \Lambda^n$ give the locations $\zeta_1, \ldots, \zeta_n$ of the n vortices. The interaction function for the point vortex model is the Hamiltonian

$$H_n(\zeta) \doteq \frac{1}{2n^2} \sum_{1 \le i < j \le n} g(\zeta_i, \zeta_j) + \frac{1}{n^2} \sum_{1 \le i \le n} h(\zeta_i). \tag{2.5}$$

Because $g(x, x')$ and $h(x)$ are not bounded continuous functions of x and x' in Λ, the point vortex model cannot be studied by the methods of this paper, but must be analyzed by other techniques [5, 8, 9, 26, 28]. The regularized point vortex model is defined like the point vortex model except that in the formula for H_n $g(x, x')$ is replaced by a suitable bounded continuous function $V(x, x')$ on Λ^2 and h is replaced by a suitable bounded continuous k on Λ.

For the regularized point vortex model the hidden space is the space $\mathcal{X}$ of probability measures on Λ while the hidden process is the sequence of empirical measures

$$Y_n(\zeta) = Y_n(\zeta, dx) \doteq \frac{1}{n} \sum_{i=1}^{n} \delta_{\zeta_i}(dx).$$

By Sanov's Theorem, this sequence satisfies the large deviation principle on $\mathcal{X}$ with rate function the relative entropy $R(\mu|\theta)$ of μ with respect to θ [10, 11, 15]. For $\mu \in \mathcal{X}$ the interaction representation function is defined by

$$\tilde{H}(\mu) \doteq \frac{1}{2} \int_{\Lambda \times \Lambda} V(x, x')\, \mu(dx)\, \mu(dx').$$

The approximation property (2.4) is easily verified.

(b) **Miller-Robert Model.** This model of the two-dimensional Euler equations is analyzed in [6], which explains in detail the physical background. For simplicity, let the flow domain be T^2, the unit torus $[0, 1) \times [0, 1)$ with periodic boundary conditions. For each $n \in I\!\!N$ let $\mathcal{L}_n$ be a uniform lattice of $a_n \doteq 2^{2n}$ sites t in T^2. The intersite spacing in each coordinate direction is 2^{-n}. Each such lattice of a_n sites induces a dyadic partition of T^2 into a_n squares called microcells, each having area $1/a_n$. For each $s \in \mathcal{L}_n$ we denote by $M(s)$ the unique microcell containing the site s in its lower left corner. The configuration spaces for the Miller-Robert model are $\Omega_n \doteq \mathcal{Y}^{a_n}$, where $\mathcal{Y}$ is a given compact subset of $I\!\!R$. Microstates are denoted by $\zeta = \{\zeta(s), s \in \mathcal{L}_n\}$. Let ρ be a probability measure on $I\!\!R$ with support $\mathcal{Y}$. P_n equals the product measure on Ω_n with identical one-dimensional marginals ρ.

There are two classes of interaction functions, the Hamiltonian and the generalized enstrophies. For $\zeta \in \Omega_n$ the Hamiltonian is defined by

$$H_{n,1}(\zeta) \doteq \frac{1}{2n^2} \sum_{s, s' \in \mathcal{L}} g_n(s - s') \zeta(s) \zeta(s'),$$

where $g_n(s - s')$ is a certain bounded continuous approximation to the lattice Green's function

$$g(s - s') \doteq \sum_{0 \neq \xi \in Z\!\!\!Z^2} |2\pi\xi|^{-2} \exp[2\pi i \langle \xi, s - s' \rangle].$$

Fix $\alpha \in I\!\!N$. For $i = 2, \ldots, \alpha + 1$ the generalized enstrophies are defined by

$$H_{n,i}(\zeta) \doteq \frac{1}{n} \sum_{s \in \mathcal{L}_n} a_i(\zeta(s)),$$

where the a_i are continuous functions mapping $\mathcal{Y}$ into $I\!\!R$.

Hypotheses 2.2 are verified in [6], to which the reader is referred for details. Let θ denote Lebesgue measure on T^2. The hidden space is the space $\mathcal{X}$ of probability measures $\mu(dx \times dy)$ on $T^2 \times \mathcal{Y}$ with first marginal θ. The hidden process is the sequence of measures

$$Y_n(dx \times dy) = Y_n(\zeta, dx \times dy) \doteq \theta(dx) \otimes \sum_{s \in \mathcal{L}_n} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy).$$

For $\mu \in \mathcal{X}$ the Hamiltonian interaction function is given by

$$\tilde{H}_1(\mu) \doteq \frac{1}{2} \int_{(T^2 \times \mathcal{Y})^2} g(x - x') y y'\, \mu(dx \times dy)\, \mu(dx' \times dy')$$

while for $i = 2, \ldots, \alpha + 1$ the interaction functions for the generalized enstrophies are given by

$$\tilde{H}_i(\mu) \doteq \int_{T^2 \times \mathcal{Y}} a_i(y)\, \mu(dx \times dy).$$

For $i = 1$ one verifies (2.4) by a detailed Fourier analysis. For $i = 2, \ldots, \alpha + 1$ (2.4) is easily verified to hold with no error term.

Given $n \in I\!\!N$ and an even integer $q < 2n$, we consider a dyadic partition of the lattice $\mathcal{L}_n$ into 2^q blocks, each block containing $a_n/2^q$ lattice sites. In correspondence with this partition we have a dyadic partition $\{D_{q,k}, k = 1, \ldots, 2^q\}$ of T^2 into macrocells. Each macrocell is the union of $a_n/2^q$ microcells $M(s)$. The large deviation principle for Y_n with respect to P_n is verified by comparing Y_n with the two-component process

$$W_{n,q}(dx \times dy) = W_{n,q}(\zeta, dx \times dy) \doteq \theta(dx) \otimes \sum_{k=1}^{2^q} 1_{D_{q,k}}(x)\, L_{n,q,k}(\zeta, dy),$$

where $L_{n,q,k}$ denotes the empirical measure $\frac{1}{a_n/2^q} \sum_{s \in D_{q,k}} \delta_{\zeta(s)}(dy)$. Through these empirical measures, $W_{n,q}$ introduces an averaging over the intermediate scale of the macrocells and thus corresponds to a coarse graining of the vorticity field. Using Sanov's Theorem, one verifies that as $n \to \infty$, $q \to \infty$, $W_{n,q}$ satisfies the two-parameter LDP on $\mathcal{X}$ with rate function the relative entropy $R(\mu|\theta \times \rho)$ of $\mu(dx \times dy)$ with respect to the product measure $\theta(dx) \times \rho(dy)$ [6, §5]. An approximation result relating Y_n and $W_{n,q}$ then allows one to prove that Y_n satisfies the LDP on $\mathcal{X}$ with the same rate function.

(c) **Quasi-geostrophic potential vorticity model.** This model of the quasi-geostrophic potential vorticity equation, described in detail in [14] and [16], incorporates the geophysical terms associated with the Coriolis effect, the deformation of an upper free surface, and bottom topography. The large deviation analysis of the model is carried out in [16].

(d) **Dispersive wave model for the nonlinear Schrödinger equation.** This model is defined in [23, 24], to which the reader is referred for details. The hidden process is a Gaussian process taking values in $L^2[0,1]$ and satisfying the LDP with respect to the prior distribution that is proved in [18]. The large deviation analysis of this model is the subject of [17]. ■

We now return to the general model. Its large deviation analysis with respect to the canonical ensemble is summarized in the next theorem. Part (a) states a variational formula for $\varphi(\beta)$, and part (b) gives the LDP for the hidden process Y_n with respect to the sequence of Gibbs measures $P_{n,\beta}$. Part (c) describes the set $\mathcal{E}_\beta$ consisting of points at which the rate function in part (b) attains its minimum of 0. Part (d) gives a concentration property of $\mathcal{E}_\beta$. As we point out after the statement of the theorem, $\mathcal{E}_\beta$ can be identified with the set of equilibrium macrostates of the statistical mechanical model. The mathematical tractability of the canonical ensemble is reflected in the simplicity of the proof of Theorem 2.4.

Theorem 2.4. *We assume Hypotheses* 2.1 *and* 2.2. *For $\beta \in I\!\!R^\sigma$ the following conclusions hold.*

(a) $\varphi(\beta) \doteq -\lim_{n\to\infty} \frac{1}{a_n} \log Z_n(a_n\beta)$ *exists and is given by*

$$\varphi(\beta) = \inf_{x\in\mathcal{X}} \{\langle \beta, \tilde{H}(x)\rangle + I(x)\}; \tag{2.6}$$

$\varphi(\beta)$ is a finite, concave, continuous function on $I\!\!R^\sigma$.

(b) *With respect to $P_{n,a_n\beta}$, Y_n satisfies the LDP on $\mathcal{X}$ with rate function*

$$I_\beta(x) \doteq I(x) + \langle \beta, \tilde{H}(x)\rangle - \inf_{y\in\mathcal{X}}\{I(y) + \langle \beta, \tilde{H}(y)\rangle\} = I(x) + \langle \beta, \tilde{H}(x)\rangle - \varphi(\beta).$$

(c) *The set $\mathcal{E}_\beta \doteq \{x \in \mathcal{X} : I_\beta(x) = 0\}$ is a nonempty, compact subset of $\mathcal{X}$. A point $\bar{x}$ lies in $\mathcal{E}_\beta$ if and only if*

$$I(\bar{x}) + \langle \beta, \tilde{H}(\bar{x})\rangle = \inf_{y\in\mathcal{X}}\{I(y) + \langle \beta, \tilde{H}(y)\rangle\} = \varphi(\beta);$$

equivalently, if and only if $\bar{x}$ solves the following unconstrained minimization problem:

$$\text{minimize } (I(x) + \langle \beta, \tilde{H}(x)\rangle) \text{ over } x \in \mathcal{X}.$$

(d) *If A is any Borel subset of $\mathcal{X}$ whose closure $\bar{A}$ satisfies $\bar{A} \cap \mathcal{E}_\beta = \emptyset$, then $I_\beta(\bar{A}) > 0$ and for some $C < \infty$*

$$P_{n,a_n\beta}\{Y_n \in A\} \le C \exp[-a_n I_\beta(\bar{A})/2] \to 0 \;\text{ as } n \to \infty.$$

Proof. (a) Since Y_n satisfies the LDP with respect to P_n, Y_n satisfies the Laplace principle with respect to P_n with the same rate function I. Hence by the approximation property

(2.4) and the boundedness and continuity of the function mapping $x \mapsto \langle \beta, \tilde{H}(x) \rangle$,

$$\begin{aligned} \varphi(\beta) &= -\lim_{n\to\infty} \frac{1}{a_n} \log Z_n(a_n \beta) \\ &= -\lim_{n\to\infty} \frac{1}{a_n} \log \int_{\Omega_n} \exp[-a_n \langle \beta, H_n \rangle] \, dP_n \\ &= -\lim_{n\to\infty} \frac{1}{a_n} \log \int_{\Omega_n} \exp[-a_n \langle \beta, \tilde{H}(Y_n) \rangle] \, dP_n \\ &= \inf_{x \in \mathcal{X}} \{ \langle \beta, \tilde{H}(x) \rangle + I(x) \}. \end{aligned}$$

This formula exhibits φ as a finite, concave function on $I\!\!R^\sigma$, which is therefore continuous on $I\!\!R^\sigma$.

(b) I_β is a rate function since I is a rate function and the function mapping $x \mapsto \langle \beta, \tilde{H}(x) \rangle$ is bounded and continuous. In order to prove that with respect to $P_{n,a_n\beta}$ Y_n satisfies the LDP with rate function I_β, it suffices to prove that with respect to $P_{n,a_n\beta}$ Y_n satisfies the Laplace principle with rate function I_β. This is an immediate consequence of (2.4) and part (a); for details, see the proof of part (b) of Theorem 3.1 in [6].

(c) $\mathcal{E}_\beta$ is a nonempty, compact subset of $\mathcal{X}$ because I_β is a rate function. The equivalent characterizations of $\bar{x} \in \mathcal{E}_\beta$ follow from the definition of I_β.

(d) If $\bar{A} \cap \mathcal{E}_\beta = \emptyset$, then for each $x \in A$ we have $I_\beta(x) > 0$. Since I_β is a rate function, it follows that $I_\beta(\bar{A}) > 0$. The large deviation upper bound in part (b) yields the display in part (d) for some $C < \infty$. The proof of the theorem is complete. ■

Part (d) of Theorem 2.4 can be regarded as a concentration property of the $P_{n,a_n\beta}$-distributions of Y_n. This property justifies calling $\mathcal{E}_\beta$ the set of equilibrium macrostates with respect to $P_{n,a_n\beta}\{Y_n \in dx\}$ or, for short, as the set of canonical equilibrium macrostates.

The next theorem further justifies the designation of $\mathcal{E}_\beta$ as the set of canonical equilibrium macrostates by relating weak limits of subsequences of $P_{n,a_n\beta}\{Y_n \in \cdot\}$ to $\mathcal{E}_\beta$. For example, if one knows that $\mathcal{E}_\beta$ consists of a unique point $\tilde{x}$, then it follows that the entire sequence $P_{n,a_n\beta}\{Y_n \in \cdot\}$ converges weakly to $\delta_{\tilde{x}}$. This situation corresponds to the absence of a phase transition. For specific models, more detailed information about weak limits of subsequences of $P_{n,a_n\beta}$ have been obtained by a number of authors including [9, 19, 26, 36].

Theorem 2.5. *We assume Hypotheses* 2.1 *and* 2.2. *For $\beta \in I\!\!R^\sigma$, any subsequence of $P_{n,a_n\beta}\{Y_n \in \cdot\}$ has a subsubsequence converging weakly to a probability measure Π_β on $\mathcal{X}$ that is concentrated on $\mathcal{E}_\beta \doteq \{x \in \mathcal{X} : I_\beta(x) = 0\}$; i.e., $\Pi_\beta\{(\mathcal{E}_\beta)^c\} = 0$. If $\mathcal{E}_\beta$ consists of a unique point $\tilde{x}$, then the entire sequence $P_{n,a_n\beta}\{Y_n \in \cdot\}$ converges weakly to $\delta_{\tilde{x}}$.*

Proof. Define $a^* \doteq \min_{n \in I\!\!N} a_n > 0$. As shown in the proof of Lemma 2.6 in [34], the large deviation upper bound given in part (b) of Theorem 2.4 implies that for each $M \in (0, \infty)$ there exists a compact subset K of $\mathcal{X}$ such that for all $n \in I\!\!N$

$$P_{n,a_n\beta}\{Y_n \in K^c\} \leq \frac{e^{-a_n M}}{1 - e^{-M}} \leq \frac{e^{-a^* M}}{1 - e^{-M}} \; .$$

It follows that the sequence $P_{n,a_n\beta}\{Y_n \in \cdot\}$ is tight and therefore that any subsequence has a subsubsequence $P_{n',a_{n'}\beta}\{Y_{n'} \in \cdot\}$ converging weakly as $n' \to \infty$ to a probability measure Π_β on $\mathcal{X}$ [Prohorov's Theorem]. In order to show that Π_β is concentrated on $\mathcal{E}_\beta$, we write the open set $(\mathcal{E}_\beta)^c$ as a union of countably many open balls V_j such that the closure $\bar{V}_j$ of each V_j has empty intersection with $\mathcal{E}_\beta$. By part (c) of Theorem 2.4 $P_{n',a_{n'}\beta}\{Y_{n'} \in V_j\} \to 0$ as $n' \to \infty$, and so

$$0 = \liminf_{n' \to \infty} P_{n',a_{n'}\beta}\{Y_{n'} \in V_j\} \geq \Pi_\beta\{V_j\}.$$

It follows that $\Pi_\beta\{V_j\} = 0$ and thus that $\Pi_\beta\{(\mathcal{E}_\beta)^c\} = 0$, as claimed.

Now assume that $\mathcal{E}_\beta = \{\tilde{x}\}$. Then the only probability measure on $\mathcal{X}$ that is concentrated on $\mathcal{E}_\beta$ is $\delta_{\tilde{x}}$. Since by the first part of the proof any subsequence of $P_{n,a_n\beta}\{Y_n \in \cdot\}$ has a subsubsequence converging weakly to $\delta_{\tilde{x}}$, it follows that the entire sequence $P_{n,a_n\beta}\{Y_n \in \cdot\}$ converges weakly to $\delta_{\tilde{x}}$. This completes the proof. ■

In the next section we consider the LDP for Y_n when conditioning is present.

3 Microcanonical Ensemble: LDP and Equilibrium Macrostates

As in the preceding section, we consider models defined in terms of a sequence of interaction functions $\{H_{n,i}, i = 1 \ldots, \sigma\}$, which are bounded measurable functions mapping Ω_n into $I\!R$. In general, the interaction functions represent conserved quantities with respect to some dynamics that underlies the model. For suitable values of $(u_1, \ldots, u_\sigma) \in I\!R^\sigma$ the ideal way to define the microcanonical ensemble is to condition the probability measure P_n on the set $\{H_{n,1} = u_1, \ldots, H_{n,\sigma} = u_\sigma\}$. However, in order to avoid problems concerning the existence of regular conditional probability distributions, we shall condition P_n on $\{H_{n,1} \in [u_1 - r, u_1 + r], \ldots, H_{n,\sigma} \in [u_\sigma - r, u_\sigma + r]\}$, where $r \in (0, 1)$. These conditioned measures, given in (3.4), define the microcanonical ensemble. Theorem 3.2 proves the LDP for the distributions of Y_n with respect to the microcanonical ensemble in the double limit obtained by sending first $n \to \infty$ and then $r \to 0$. We then define, in terms of the rate function in this LDP, the set of microcanonical equilibrium macrostates and derive some of its properties.

For $u = (u_1, \ldots, u_\sigma) \in I\!R^\sigma$ a key role in the large deviation analysis of the microcanonical ensemble is played by

$$J(u) \doteq \inf\{I(x) : x \in \mathcal{X}, \tilde{H}(x) = u\}. \tag{3.1}$$

In terms of J the canonical free energy $\varphi(\beta)$, given in part (a) of Theorem 2.4 by

$$\varphi(\beta) = \inf_{x \in \mathcal{X}} \{\langle \beta, \tilde{H}(x) \rangle + I(x)\},$$

can be rewritten as

$$\begin{aligned}\varphi(\beta) &= \inf_{u\in I\!\!R^\sigma}\Big\{\inf\{\langle\beta,\tilde H(x)\rangle + I(x) : x\in\mathcal{X}, \tilde H(x)=u\}\Big\}\\ &= \inf_{u\in I\!\!R^\sigma}\{\langle\beta,u\rangle + J(u)\}.\end{aligned}$$

Introducing the microcanonical entropy

$$s(u) \doteq -J(u) = -\inf\{I(x) : x\in\mathcal{X}, \tilde H(x) = u\}, \tag{3.2}$$

we have

$$\varphi(\beta) = \inf_{u\in I\!\!R^\sigma}\{\langle\beta,u\rangle - s(u)\}. \tag{3.3}$$

This formula expresses φ as the Legendre-Fenchel transform of s. The microcanonical entropy will play a central role in the results on equivalence and nonequivalence of the canonical and microcanonical ensembles to be presented in Section 4.

The function J plays other roles in the theory. Since each $\tilde H_i$ is a bounded continuous function mapping $\mathcal{X}$ into $I\!\!R$ and since with respect to P_n Y_n satisfies the LDP on $\mathcal{X}$ with rate function I, it follows from the contraction principle that with respect to P_n $\tilde H(Y_n) = (\tilde H_1(Y_n),\ldots,\tilde H_\sigma(Y_n))$ satisfies the LDP on $I\!\!R^\sigma$ with rate function J [10, Thm. 4.2.1]. When expressed in terms of the equivalent Laplace principle, this means that for any bounded continuous function g mapping $I\!\!R^\sigma$ into $I\!\!R$

$$\lim_{n\to\infty}\frac{1}{a_n}\log\int_{\Omega_n}\exp[a_n\, g(\tilde H(Y_n))]\, dP_n = \sup_{u\in I\!\!R^\sigma}\{g(u) - J(u)\}.$$

Because of the approximation property (2.4), this readily extends to the Laplace principle on $I\!\!R^\sigma$, and thus the LDP on $I\!\!R^\sigma$, for $H_n \doteq (H_{n,1},\ldots,H_{n,\sigma})$.

In part (a) of the next proposition we record the LDP's just discussed and two properties of the microcanonical entropy. When applied to the regularized point vortex model, the LDP for the P_n-distributions of H_n generalizes the large deviation estimates obtained in [19, Thm. 2.1]. In parts (b) and (c) of the proposition some related facts needed later in this section are given. We define $\operatorname{dom} J$ to be the set of $u\in I\!\!R^\sigma$ for which $J(u)<\infty$. For $r\in(0,1)$ and $u\in\operatorname{dom} J$, we also define

$$\{u\}^{(r)} \doteq [u_1 - r, u_1 + r]\times\cdots\times[u_\sigma - r, u_\sigma + r].$$

Part (b) is a consequence of the LDP for H_n given in part (a) and of the bound $J(\mathrm{int}(\{u\}^{(r)})) \leq J(u)$. Part (c) follows from the lower semicontinuity of J and from part (b).

Proposition 3.1. *We assume Hypotheses* 2.1 *and* 2.2. *The following conclusions hold.*

(a) *With respect to P_n, the sequences $\tilde H(Y_n)$ and H_n satisfy the LDP on $I\!\!R^\sigma$ with rate function J. Hence $s \doteq -J$ is nonpositive and upper semicontinuous.*

(b) *For $u\in\operatorname{dom} J$ and any $r\in(0,1)$*

$$\begin{aligned}-J(u) &\leq \liminf_{n\to\infty}\frac{1}{a_n}\log P_n\{H_n\in\{u\}^{(r)}\}\\ &\leq \limsup_{n\to\infty}\frac{1}{a_n}\log P_n\{H_n\in\{u\}^{(r)}\} \;\leq\; -J(\{u\}^{(r)}).\end{aligned}$$

(c) *As* $r \to 0$, $J(\{u\}^{(r)}) \nearrow J(u)$. *Hence*

$$\lim_{r\to 0}\lim_{n\to\infty}\frac{1}{a_n}\log P_n\{H_n \in \{u\}^{(r)}\} = -J(u).$$

The main theorem of this section is the LDP for Y_n with respect to the microcanonical ensemble, given in Theorem 3.2. For $A \in \mathcal{F}_n$ this ensemble is defined by the conditioned measures

$$P_n^{u,r}\{A\} \doteq P_n\{A \,|\, H_n \in \{u\}^{(r)}\}, \tag{3.4}$$

where $u \in \operatorname{dom} J$ and $r \in (0,1)$. For all sufficiently large n it follows from part (b) of Proposition 3.1 that $P_n\{H_n \in \{u\}^{(r)}\} > 0$ and hence that $P_n^{u,r}$ is well defined.

Theorem 3.2. *Take* $u \in \operatorname{dom} J$ *and assume Hypotheses* 2.1 *and* 2.2. *With respect to the conditioned measures* $P_n^{u,r}$, Y_n *satisfies the LDP on* $\mathcal{X}$, *in the double limit* $n \to \infty$ *and* $r \to 0$, *with rate function*

$$I^u(x) \doteq \begin{cases} I(x) - J(u) & \text{if } \tilde{H}(x) = u, \\ \infty & \text{otherwise.} \end{cases}$$

That is, for any closed subset F *of* $\mathcal{X}$

$$\lim_{r\to 0}\limsup_{n\to\infty}\frac{1}{a_n}\log P_n^{u,r}\{Y_n \in F\} \le -I^u(F) \tag{3.5}$$

and for any open subset G *of* $\mathcal{X}$

$$\lim_{r\to 0}\liminf_{n\to\infty}\frac{1}{a_n}\log P_n^{u,r}\{Y_n \in G\} \ge -I^u(G). \tag{3.6}$$

We first prove that I^u defines a rate function. Clearly I^u is nonnegative. For $u \in \operatorname{dom} J$ and $M < \infty$

$$\{x \in \mathcal{X} : I^u(x) \le M\} = \{x \in \mathcal{X} : I(x) \le M + J(u)\} \cap \tilde{H}^{-1}(\{u\}).$$

Since $J(u) < \infty$, I has compact level sets, and $\tilde{H}^{-1}(\{u\})$ is closed, it follows that I^u has compact level sets.

Concerning the large deviation bounds in Theorem 3.2, we offer two proofs. The first is preferred because it is close to the heuristic sketch of the LDP given in the introduction. Throughout the two proofs we fix $u \in \operatorname{dom} J$.

The first proof of the large deviation upper bound actually derives a stronger inequality. Namely, for all sufficiently small $r \in (0,1)$ and any closed subset F of $\mathcal{X}$

$$\limsup_{n\to\infty}\frac{1}{a_n}\log P_n^{u,r}\{Y_n \in F\} \le -I^u(F). \tag{3.7}$$

For any $x \in \mathcal{X}$ and $\alpha > 0$ we denote by $\bar{B}(x,\alpha)$ and $B(x,\alpha)$ the closed ball and the open ball in $\mathcal{X}$ with center x and radius α. Let $\delta > 0$ be given. Since I is lower semicontinuous, for any $x \in \mathcal{X}$ and all sufficiently small $\alpha > 0$ we have $I(\bar{B}(x,\alpha)) \geq I(x) - \delta$. Now take any $x \in \mathcal{X}$ such that $\tilde{H}(x) = u$. For any $r \in (0,1)$ and all sufficiently small α the large deviation upper bound for Y_n with respect to P_n and part (b) of Proposition 3.1 yield

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in \bar{B}(x,\alpha)\} \qquad (3.8)\\
&\quad \leq \limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{\{Y_n \in \bar{B}(x,\alpha)\} \cap \{H_n \in \{u\}^{(r)}\}\} \\
&\qquad\quad - \liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in \{u\}^{(r)}\} \\
&\quad \leq \limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{Y_n \in \bar{B}(x,\alpha)\} \\
&\qquad\quad - \liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in \{u\}^{(r)}\} \\
&\quad \leq -I(\bar{B}(x,\alpha)) + J(u) \\
&\quad \leq -I(x) + J(u) + \delta \\
&\quad = -I^u(x) + \delta.
\end{aligned}$$

Now take any $x \in \mathcal{X}$ such that $\tilde{H}(x) \neq u$. Thus $I^u(x) = \infty$, and there exists $t \in (0,1)$ such that $\tilde{H}(x) \notin \{u\}^{(t)}$. By the approximation property (2.4) and the continuity of $\tilde{H}$, for any $r \in (0,t)$, all sufficiently small $\alpha > 0$, and all sufficiently large n we have

$$\{Y_n \in \bar{B}(x,\alpha)\} \cap \{H_n \in \{u\}^{(r)}\} \subset \{Y_n \in \bar{B}(x,\alpha)\} \cap \{\tilde{H}(Y_n) \in \{u\}^{(t)}\} = \emptyset.$$

Hence for such r and α

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in \bar{B}(x,\alpha)\} \\
&\quad \leq \limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{\{Y_n \in \bar{B}(x,\alpha)\} \cap \{H_n \in \{u\}^{(r)}\}\} \\
&\qquad\quad - \liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in \{u\}^{(r)}\} \\
&\quad = -\infty = -I^u(x).
\end{aligned}$$

We have proved that for any $x \in \mathcal{X}$, all sufficiently small $r \in (0,1)$, and all sufficiently small $\alpha > 0$

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in \bar{B}(x,\alpha)\} \leq -I^u(x) + \delta.$$

Let F be a compact subset of $\mathcal{X}$. We can cover F with finitely many closed balls $\bar{B}(x_i,\alpha_i)$ with $x_i \in F$ and $\alpha_i > 0$ so small that the last display is valid for $x = x_i$, all sufficiently small $r \in (0,1)$, and $\alpha = \alpha_i$. It follows that for all sufficiently small $r \in (0,1)$

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in F\} \leq -\min_i I^u(x_i) + \delta \leq -I(F) + \delta.$$

Sending $\delta \to 0$ yields the upper bound (3.7). Finally, for any closed set F the upper bound (3.7) is a consequence of the following uniform exponential tightness estimate.

Lemma 3.3. *Fix $u \in \operatorname{dom} J$. Then for all sufficiently large $M \in (0,\infty)$ there exists a compact subset D of $\mathcal{X}$ such that for every $r \in (0,1)$*

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in D^c\} \le -M.$$

Proof. Given $u \in \operatorname{dom} J$, we take $M > J(u)$. As shown in the proof of Lemma 2.6 in [34], the large deviation upper bound satisfied by Y_n with respect to P_n implies that there exists a compact subset D of $\mathcal{X}$ such that

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{Y_n \in D^c\} \le -2M.$$

Since for every $r \in (0,1)$

$$P_n^{u,r}\{Y_n \in D^c\} \le \frac{P_n\{Y_n \in D^c\}}{P_n\{H_n \in \{u\}^{(r)}\}},$$

it follows from part (b) of Proposition 3.1 that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in D^c\} \\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{Y_n \in D^c\} - \liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in \{u\}^{(r)}\} \\
&\quad \le -2M + J(u) \le -M.
\end{aligned}$$

This completes the proof. ■

We next prove the large deviation lower bound in Theorem 3.2 by showing that for any fixed $r \in (0,1)$ and any open subset G of $\mathcal{X}$

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in G\} \ge -I^u(G) + J(\{u\}^{(r)}) - J(u). \tag{3.9}$$

Sending $r \to 0$ and using part (c) of Proposition 3.1 yields the large deviation lower bound in Theorem 3.2.

Let x be any point in G such that $\tilde{H}(x) = u$. By the approximation property (2.4) and the continuity of $\tilde{H}$, for any number r^- satisfying $0 < r^- < r$ and all sufficiently large n, we can choose $\alpha > 0$ to be so small that $B(x,\alpha) \subset G$ and

$$\begin{aligned}
\{Y_n \in B(x,\alpha)\} \cap \{H_n \in \{u\}^{(r)}\} &\supset \{Y_n \in B(x,\alpha)\} \cap \{\tilde{H}(Y_n) \in \{u\}^{(r^-)}\} \\
&= \{Y_n \in B(x,\alpha)\}.
\end{aligned}$$

Hence for such α, the large deviation lower bound for Y_n with respect to P_n and part (b) of Proposition 3.1 yield

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in G\}$$

$$
\begin{aligned}
&\geq \liminf_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in B(x,\alpha)\} \\
&\geq \liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{\{Y_n \in B(x,\alpha)\} \cap \{H_n \in \{u\}^{(r)}\}\} \\
&\qquad - \limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in \{u\}^{(r)}\} \\
&\geq \liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{Y_n \in B(x,\alpha)\} \\
&\qquad - \limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{H_n \in \{u\}^{(r)}\} \\
&\geq -I(B(x,\alpha)) + J(\{u\}^{(r)}) \\
&\geq -I(x) + J(\{u\}^{(r)}) \\
&= -I^u(x) + J(\{u\}^{(r)}) - J(u).
\end{aligned}
$$

Now take any $x \in \mathcal{X}$ such that $\tilde{H}(x) \neq u$. Since $I^u(x) = \infty$, it follows that

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in G\} \geq -\infty = -I^u(x) + J(\{u\}^{(r)}) - J(u).$$

We have thus obtained the same lower bound for all $x \in G$. We conclude that

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{1}{a_n} \log P_n^{u,r}\{Y_n \in G\} \quad &\geq \quad \sup_{x\in G}\{-I^u(x)\} + J(\{u\}^{(r)}) - J(u) \\
&= \quad -I^u(G) + J(\{u\}^{(r)}) - J(u).
\end{aligned}
$$

This completes the proof of the large deviation lower bound (3.9). The proof of Theorem 3.2 is done.

The second proof of the large deviation bounds in Theorem 3.2 uses the following alternate representation for the rate function:

$$I^u(x) = I(\{x\} \cap \tilde{H}^{-1}(\{u\})).$$

Let F be any closed subset of $\mathcal{X}$. We choose ψ to be any function mapping $(0,1)$ onto $(0,1)$ with the properties that $\psi(r) > r$ for all $r \in (0,1)$ and $\lim_{r\to 0} \psi(r) = 0$. Clearly, as $r \downarrow 0$, $\{u\}^{(\psi(r))} \downarrow \{u\}$. We need the limit

$$\lim_{r\to 0} I(F \cap \tilde{H}^{-1}(\{u\}^{(\psi(r))})) = I(F \cap \tilde{H}^{-1}(u)),$$

which follows from routine calculations using the continuity of $\tilde{H}$ and the fact that I^u is a rate function. The proof of this limit is omitted. The rest of the proof of the large deviation upper bound is straightforward. By the approximation property (2.4) and the continuity of $\tilde{H}$, for any $r \in (0,1)$ and all sufficiently large n

$$P_n\{\{Y_n \in F\} \cap \{H_n \in \{u\}^{(r)}\}\} \leq P_n\{\{Y_n \in F\} \cap \{\tilde{H}(Y_n) \in \{u\}^{(\psi(r))}\}\}.$$

Then the large deviation upper bound for Y_n with respect to P_n and part (c) of Proposition 3.1 yield

$$\begin{aligned}
&\lim_{r\to 0}\limsup_{n\to\infty}\frac{1}{a_n}\log P_n^{u,r}\{Y_n\in F\}\\
&\quad\le \lim_{r\to 0}\limsup_{n\to\infty}\frac{1}{a_n}\log P_n\{Y_n\in[F\cap\tilde{H}^{-1}(\{u\}^{(\psi(r))})]\}\\
&\qquad -\lim_{r\to 0}\liminf_{n\to\infty}\frac{1}{a_n}\log P_n\{H_n\in\{u\}^{(r)}\}\\
&\quad\le -\lim_{r\to 0} I(F\cap\tilde{H}^{-1}(\{u\}^{(\psi(r))}))+J(u)\\
&\quad= -I(F\cap\tilde{H}^{-1}(u))+J(u)\\
&\quad= -I^u(F).
\end{aligned}$$

This is the large deviation upper bound (3.5).

Now let G be any open subset of $\mathcal{X}$. Again by the approximation property (2.4) and the continuity of $\tilde{H}$, for any number r^- satisfying $0<r^-<r$ and all sufficiently large n

$$\begin{aligned}
&P_n\{\{Y_n\in G\}\cap\{H_n\in\{u\}^{(r)}\}\}\\
&\quad\ge P_n\{\{Y_n\in G\}\cap\{\tilde{H}(Y_n)\in\{u\}^{(r^-)}\}\}\\
&\quad\ge P_n\{Y_n\in G\cap\tilde{H}^{-1}(\mathrm{int}\{u\}^{(r^-)})\}.
\end{aligned}$$

The large deviation lower bound for Y_n with respect to P_n and part (c) of Proposition 3.1 yield

$$\begin{aligned}
&\lim_{r\to 0}\liminf_{n\to\infty}\frac{1}{a_n}\log P_n^{u,r}\{Y_n\in G\}\\
&\quad\ge \lim_{r\to 0}\liminf_{n\to\infty}\frac{1}{a_n}\log P_n\{Y_n\in[G\cap\tilde{H}^{-1}(\mathrm{int}\{u\}^{(r^-)})]\}\\
&\qquad -\lim_{r\to 0}\lim_{n\to\infty}\frac{1}{a_n}\log P_n\{H_n\in\{u\}^{(r)}\}\\
&\quad\ge -\lim_{r\to 0} I(G\cap\tilde{H}^{-1}(\mathrm{int}\{u\}^{(r^-)}))+J(u)\\
&\quad\ge -I(G\cap\tilde{H}^{-1}(u))+J(u)\\
&\quad= -I^u(G).
\end{aligned}$$

This is the large deviation lower bound (3.6), completing the second proof of the large deviation bounds in Theorem 3.2. The proof of Theorem 3.2 is done.

In Section 2 the large deviation analysis of the canonical ensemble led us to define, in terms of the rate function in the corresponding LDP, the set of canonical equilibrium macrostates. Analogously, for $u\in\operatorname{dom}J$ we define, in terms of the rate function I^u in Theorem 3.2, the set of microcanonical equilibrium macrostates

$$\mathcal{E}^u \doteq \{x\in\mathcal{X}: I^u(x)=0\}.$$

Thus $\bar{x} \in \mathcal{E}^u$ if and only if $I(\bar{x}) = J(u)$ and $\tilde{H}(\bar{x}) = u$. We next point out that in certain models elements of $\mathcal{E}^u$ have an equivalent characterization in terms of constrained maximum entropy principles.

Remark 3.4. Equivalent characterization in terms of constrained maximum entropy principles. Since $J(u)$ equals the infimum of I over all elements x satisfying the constraint $\tilde{H}(x) = u$, we see that $\bar{x} \in \mathcal{E}^u$ if and only if $\bar{x}$ solves the following constrained minimization problem:

$$\text{minimize } I(x) \text{ over } x \in \mathcal{X} \text{ subject to the constraint } \tilde{H}(x) = u.$$

Both for the regulariued point vortex model and the Miller-Robert model the rate function I equals a relative entropy, which in turn equals minus the physical entropy. Hence for these models the last display gives an equivalent characteriuation of microcanonical equilibrium macrostates in terms of a constrained maximum entropy principle. ■

Parts (c) and (d) of Theorem 2.4 state several properties of the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates. The next theorem gives analogous properties of $\mathcal{E}^u$. The second of these properties is slightly more complicated than in the canonical case because the microcanonical measures $P_n^{u,r}$ depend on the two parameters $n \in I\!N$ and $r \in (0,1)$.

Theorem 3.5. *We assume Hypotheses* 2.1 *and* 2.2. *For* $u \in \operatorname{dom} J$ *the following conclusions hold.*

(a) $\mathcal{E}^u \doteq \{x \in \mathcal{X} : I^u(x) = 0\}$ *is a nonempty, compact subset of* $\mathcal{X}$. *A point* $\bar{x} \in \mathcal{X}$ *lies in* $\mathcal{E}^u$ *if and only if* $I(\bar{x}) = J(u)$ *and* $\tilde{H}(\bar{x}) = u$; *equivalently, if and only if* $\bar{x}$ *solves the following constrained minimization problem:*

$$\textit{minimize } I(x) \textit{ over } x \in \mathcal{X} \textit{ subject to the constraint } \tilde{H}(x) = u.$$

(b) *Let* A *be any Borel subset of* $\mathcal{X}$ *whose closure* $\bar{A}$ *satisfies* $\bar{A} \cap \mathcal{E}^u = \emptyset$. *Then* $I^u(\bar{A}) > 0$. *In addition, there exists* $r_0 \in (0,1)$ *and for all* $r \in (0, r_0]$ *there exists* $C_r < \infty$ *such that*

$$P_n^{u,r}\{Y_n \in A\} \leq C_r \exp[-a_n I^u(\bar{A})/2] \to 0 \;\; \textit{as } n \to \infty.$$

Proof. (a) $\mathcal{E}^u$ is a nonempty, compact subset of $\mathcal{X}$ because I^u is a rate function. The equivalent characterizations of $\bar{x} \in \mathcal{E}^u$ follow from the formula for I^u.

(b) If $\bar{A} \cap \mathcal{E}^u = \emptyset$, then for each $x \in A$ we have $I^u(x) > 0$. Since I^u is a rate function, it follows that $I^u(\bar{A}) > 0$. The large deviation upper bound for the $P_n^{u,r}$-distributions of Y_n given in (3.5) completes the proof. ■

Part (b) of Theorem 3.5 can be regarded as a concentration property of the $P_n^{u,r}$-distributions of Y_n. This property justifies calling $\mathcal{E}^u$ the set of microcanonical equilibrium macrostates.

Theorem 2.5 studies compactness properties of the sequence of $P_{n,a_n\beta}$-distributions of Y_n and shows that any weak limit of a convergent subsequence of this sequence is concentrated on $\mathcal{E}_\beta$. In the next theorem we formulate an analogue for the microcanonical ensemble, studying compactness and weak limit properties of the $P_n^{u,r}$-distributions of Y_n. In the case of the classical lattice gas, a related result is given, for example, in [12, Lem. 4.1].

Theorem 3.6. *We assume Hypotheses* 2.1 *and* 2.2. *For* $u \in \operatorname{dom} J$ *the following conclusions hold.*

(a) *For* $r \in (0,1)$, *any subsequence of* $P_n^{u,r}\{Y_n \in \cdot\}$ *has a subsubsequence* $P_{n'}^{u,r}\{Y_{n'} \in \cdot\}$ *converging weakly to a probability measure* $\Pi^{u,r}$ *on* $\mathcal{X}$ *as* $n' \to \infty$.

(b) *There exists* $r_0 \in (0,1)$ *such that for all* $r \in (0, r_0]$ $\Pi^{u,r}$ *is concentrated on* $\mathcal{E}^u$; *i.e.,* $\Pi^{u,r}\{(\mathcal{E}^u)^c\} = 0$. *Thus if* $\mathcal{E}^u$ *consists of a unique point* $\tilde{x}$, *then for all* $r \in (0, r_0]$ *the entire sequence* $P_n^{u,r}\{Y_n \in \cdot\}$ *converges weakly to* $\delta_{\tilde{x}}$ *as* $n \to \infty$.

(c) *For any sequence* $r_k \subset (0,1)$ *converging to* 0, *any subsequence of* Π^{u,r_k} *has a subsubsequence converging weakly to a probability measure* Π^u *on* $\mathcal{X}$ *that is concentrated on* $\mathcal{E}^u$.

Proof. (a) Define $a^* \doteq \min_{n \in I\!N} a_n > 0$. The exponential tightness estimate in Lemma 3.3 implies that for all sufficiently large $M \in (0,\infty)$ there exists a compact subset D of $\mathcal{X}$ such that for all $r \in (0,1)$ and all sufficiently large n

$$P_n^{u,r}\{Y_n \in D^c\} \leq \exp[-a_n M/2] \leq \exp[-a^* M/2]. \tag{3.10}$$

Since M can be taken to be arbitrarily large, this yields the tightness of the sequence $P_n^{u,r}\{Y_n \in \cdot\}$. The tightness implies that any subsequence of $P_n^{u,r}\{Y_n \in \cdot\}$ has a subsubsequence $P_{n'}^{u,r}\{Y_{n'} \in \cdot\}$ converging weakly to a probability measure $\Pi^{u,r}$ on $\mathcal{X}$ as $n' \to \infty$ [Prohorov's Theorem]. This completes the proof of part (a).

(b) We use the value of r_0 from part (b) of Theorem 3.5. As in the proof of Theorem 2.5, in order to prove the concentration property of $\Pi^{u,r}$, we write the open set $(\mathcal{E}^u)^c$ as a union of countably many open balls V_j such that the closure $\bar{V}_j$ of each V_j has empty intersection with $\mathcal{E}^u$. Let $P_{n'}^{u,r}\{Y_{n'} \in \cdot\} \Rightarrow \Pi^{u,r}$ be the subsubsequence arising in the proof of part (a) of the present theorem. For $r \in (0, r_0]$, part (b) of Theorem 3.5 implies that $P_{n'}^{u,r}\{Y_{n'} \in V_j\} \to 0$ as $n' \to \infty$, and so

$$0 = \liminf_{n' \to \infty} P_{n'}^{u,r}\{Y_{n'} \in V_j\} \geq \Pi^{u,r}\{V_j\}.$$

It follows that $\Pi^{u,r}\{V_j\} = 0$ and thus that $\Pi^{u,r}\{(\mathcal{E}^u)^c\} = 0$, as claimed. If $\mathcal{E}^u$ consists of a unique point $\tilde{x}$, then as in the proof of Theorem 2.5, one shows that as $n \to \infty$ $P_n^{u,r}\{Y_n \in \cdot\} \Rightarrow \delta_{\tilde{x}}$. This completes the proof of part (b).

(c) This follows from part (b), Prohorov's Theorem, and the compactness of $\mathcal{E}^u$. The proof of Theorem 3.6 is complete. ■

4 Equivalence and Nonequivalence of Ensembles

In the preceding section we presented, for the microcanonical ensemble, analogues of results proved for the canonical ensemble in Section 2. These include large deviation theorems and properties of the set of equilibrium macrostates. Such analogues of results for the two ensembles point to a much deeper relationship between them. As we will soon see, the two ensembles are intimately related both at the level of thermodynamic functions and at the level of equilibrium macrostates, and the results at these two levels mirror each other.

Our main results on equivalence and nonequivalence of ensembles at the level of equilibrium macrostates are presented in Theorems 4.4, 4.6, and 4.8 and are summarized in Figure 1. Definitive and complete, they express, in terms of concavity and other properties of the microcanonical entropy, relationships between the sets of canonical and microcanonical equilibrium macrostates. The proofs of these relationships are based on straightforward concave analysis. Other results in this section explore related issues. For example, Corollary 4.7 is a uniqueness result for equilibrium macrostates, Theorem 4.10 relates the equivalence of ensembles to the differentiability of the canonical free energy, and Theorem 4.11 shows that a certain equivalence-of-ensemble relationship implies a concavity property of the microcanonical entropy.

We start our presentation by recalling an elementary result at the level of thermodynamic functions. The microcanonical entropy is the nonpositive function defined for $u \in I\!R^\sigma$ by

$$s(u) \doteq -J(u) \doteq -\inf\{I(x) : x \in \mathcal{X}, \tilde{H}(x) = u\}.$$

We define $\operatorname{dom} s$ as the set of $u \in I\!R^\sigma$ for which $s(u) > -\infty$. As shown in (3.3), the canonical free energy $\varphi(\beta)$ can be obtained from s by the formula

$$\varphi(\beta) = \inf_{u \in I\!R^\sigma}\{\langle \beta, u\rangle - s(u)\}, \tag{4.1}$$

which expresses φ as the Legendre-Fenchel transform s^* of s. In general, $\varphi = s^*$ is finite, concave, and continuous on $I\!R^\sigma$ [Thm. 2.4(a)], and s is upper semicontinuous [Prop. 3.1(a)]. If it is the case that s is concave on $I\!R^\sigma$, then concave function theory implies that s equals the Legendre-Fenchel transform of φ [45, p. 104]; viz., for $u \in I\!R^\sigma$

$$s(u) = \varphi^*(u) = \inf_{\beta \in I\!R^\sigma}\{\langle \beta, u\rangle - \varphi(\beta)\}. \tag{4.2}$$

If s is concave on $I\!R^\sigma$, then following standard terminology in the statistical mechanical literature, we say that the canonical ensemble and the microcanonical ensemble are thermodynamically equivalent [28, 33]. As we will see, when properly interpreted, the nonconcavity of s at points $u \in I\!R^\sigma$ will imply that the ensembles are nonequivalent at the level of equilibrium macrostates for those values of u [Thm. 4.4(b)]. Further connections between thermodynamic equivalence of ensembles and equivalence of ensembles at the level of equilibrium macrostates are made explicit in Theorem 4.9. In particular, under a hypothesis on the domains of various functions that is not necessarily satisfied in all models of interest, thermodynamic equivalence of ensembles is a necessary and sufficient condition for equivalence of ensembles to hold at the level of equilibrium macrostates [Thm. 4.9(c)].

The concavity of s on $I\!R^\sigma$ depends on the nature of I and $\tilde{H}$. For example, if I is concave on $\mathcal{X}$ and $\tilde{H}$ is affine, then s is concave on $I\!R^\sigma$. However, in general the concavity of s is not valid. In fact, because of the local mean-field, long-range nature of the Hamiltonians arising in many models of turbulence, including the Miller-Robert model [Example 2.3(b)], the associated microcanonical entropies are typically not concave on subsets of $I\!R^\sigma$ corresponding to a range of negative temperatures.

In order to see how concavity properties of s determine relationships between the sets of equilibrium macrostates, we define for $u \in I\!R^\sigma$ the concave function

$$s^{**}(u) \doteq \inf_{\beta \in I\!R^\sigma}\{\langle \beta, u\rangle - s^*(\beta)\} = \inf_{\beta \in I\!R^\sigma}\{\langle \beta, u\rangle - \varphi(\beta)\}.$$

Because of (4.2), it is obvious that s is concave on $I\!R^\sigma$ if and only if s and s^{**} coincide. Whenever $s(u) > -\infty$ and $s(u) = s^{**}(u)$, we shall say that s is concave at u.

Now assume that s is not concave on $I\!R^\sigma$. Since for any $u \in \operatorname{dom} s$ and all $\beta \in I\!R^\sigma$ we have $s(u) \leq \langle \beta, u\rangle - s^*(\beta)$, it follows that for all $u \in I\!R^\sigma$

$$s(u) \leq \inf_{\beta \in I\!R^\sigma}\{\langle \beta, u\rangle - s^*(\beta)\} = s^{**}(u). \tag{4.3}$$

In addition, if f is any upper semicontinuous, concave function satisfying $s(u) \leq f(u)$ for all $u \in I\!R^\sigma$, then for all $\beta \in I\!R^\sigma$ $s^*(\beta) \geq f^*(\beta)$ and thus $s^{**}(u) \leq f^{**}(u) = f(u)$ for all $u \in I\!R^\sigma$. It follows that if s is not concave on $I\!R^\sigma$, then s^{**} is the upper semicontinuous, concave hull of s; i.e., the smallest upper semicontinuous, concave function on $I\!R^\sigma$ that majorizes s. In particular, if $s(u) > -\infty$, then $s^{**}(u) > -\infty$; thus $\operatorname{dom} s \subset \operatorname{dom} s^{**}$.

Since s^{**} is an upper semicontinuous, concave function, we can introduce a basic concept in concave function theory that will play a key role in our results on equivalence and nonequivalence of ensembles. For $u \in \operatorname{dom} s^{**}$ the superdifferential of s^{**} at u is defined as the set $\partial s^{**}(u)$ consisting of $\beta \in I\!R^\sigma$ such that

$$s^{**}(w) \leq s^{**}(u) + \langle \beta, w - u\rangle \text{ for all } w \in I\!R^\sigma; \tag{4.4}$$

any such β is called a supergradient of s^{**} at u. The effective domain of the superdifferential of s^{**} is defined to be the set $\operatorname{dom} \partial s^{**}$ consisting of $u \in I\!R^\sigma$ for which $\partial s^{**}(u)$ is nonempty. It can be shown that [45, p. 217]

$$\operatorname{ri}(\operatorname{dom} s^{**}) \subset \operatorname{dom} \partial s^{**} \subset \operatorname{dom} s^{**}, \tag{4.5}$$

where for A a subset of $I\!R^\sigma$ $\operatorname{ri}(\operatorname{dom} A)$ denotes the relative interior of A. These relationships imply that $\partial s^{**}(u)$ is nonempty for $u \in \operatorname{dom} s^{**}$ except possibly for u in the relative boundary of $\operatorname{dom} s^{**}$.

The purpose of this section is to investigate, in terms of concavity properties of s and s^{**}, relationships between the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates and the set $\mathcal{E}^u$

of microcanonical equilibrium macrostates. We recall that for $\beta \in I\!\!R^\sigma$ and $u \in \mathrm{dom}\, s$ these sets are defined by

$$\begin{aligned}\mathcal{E}_\beta &= \{x \in \mathcal{X} : I_\beta(x) = 0\} \\ &= \left\{x \in \mathcal{X} : I(x) + \langle \beta, \tilde{H}(x)\rangle = \inf_{y \in \mathcal{X}}\{I(y) + \langle \beta, \tilde{H}(y)\rangle\} = \varphi(\beta)\right\}\end{aligned}$$

and

$$\mathcal{E}^u \doteq \{x \in \mathcal{X} : I^u(x) = 0\} = \{x \in \mathcal{X} : \tilde{H}(x) = u, I(x) = -s(u)\}.$$

I_β is the rate function in the LDP for the canonical ensemble [Thm. 2.4], and I^u is the rate function in the LDP for the microcanonical ensemble [Thm. 3.2]. As the sets of points at which the corresponding rate functions attain their minimum of 0, both $\mathcal{E}_\beta$ for $\beta \in I\!\!R^\sigma$ and $\mathcal{E}^u$ for $u \in \mathrm{dom}\, s$ are nonempty and compact. It is convenient to extend the definition of $\mathcal{E}^u$ to all $u \in I\!\!R^\sigma$ by defining $\mathcal{E}^u = \emptyset$ for $u \in I\!\!R^\sigma \setminus \mathrm{dom}\, s$.

First-order differentiability conditions show that relationships between $\mathcal{E}_\beta$ and $\mathcal{E}^u$ are plausible. In fact, the first-order condition for $x^* \in \mathcal{X}$ to be in $\mathcal{E}_\beta$ is

$$I'(x^*) + \langle \beta, \tilde{H}'(x^*)\rangle = 0, \tag{4.6}$$

where $'$ denotes the Frechet derivative and we assume that I and $\tilde{H}$ are Frechet-differentiable. The first-order condition for $x^* \in \mathcal{X}$ to be in $\mathcal{E}^u$ is also (4.6), where β is a Lagrange multiplier dual to the constraint $\tilde{H}(x^*) = u$. In order to see the precise relationships between $\mathcal{E}^u$ and $\mathcal{E}_\beta$, we need a more detailed analysis.

As we will see, there are three possible relationships that can occur between $\mathcal{E}^u$ and $\mathcal{E}_\beta$. If for a given $u \in \mathrm{dom}\, s$ there exists $\beta \in I\!\!R^\sigma$ such that $\mathcal{E}^u = \mathcal{E}_\beta$, then the ensembles are said to be fully equivalent or that full equivalence of ensembles holds. If instead of equality $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$ for some $\beta \in I\!\!R^\sigma$, then the ensembles are said to be partially equivalent or that partial equivalence of ensembles holds. It may also happen that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \in I\!\!R^\sigma$. If this occurs, then the microcanonical ensemble is said to be nonequivalent to any canonical ensemble or that nonequivalence of ensembles holds. It is convenient to group the first two cases together. If for a given u there exists β such that either $\mathcal{E}^u$ equals $\mathcal{E}_\beta$ or $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$, then the ensembles are said to be equivalent or that equivalence of ensembles holds.

The probabilistic role played by $\mathcal{E}^u$ and $\mathcal{E}_\beta$ should be kept in mind when interpreting these relationships. According to part (c) of Theorem 2.4, for any Borel subset A whose closure is disjoint from $\mathcal{E}_\beta$, $P_{n,a_n\beta}\{Y_n \in A\} \to 0$. Theorem 2.5 refines this by showing that convergent subsequences of $P_{n,a_n\beta}\{Y_n \in \cdot\}$ have weak limits with support in $\mathcal{E}_\beta$. Theorems 3.5 and 3.6 do the same for the microcanonical ensemble. Only when $\mathcal{E}_\beta = \mathcal{E}^u = \{x\}$ can we be sure that the two ensembles give the same prediction in the sense of weak convergence. A condition implying these equalities is given in Corollary 4.7.

A key insight revealed by our results is that the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates can be richer than the set $\mathcal{E}_\beta$ of canonical equilibrium macrostates. Specifically, every $x \in \mathcal{E}_\beta$ is also in $\mathcal{E}^u$ for some u, but if the microcanonical entropy s is not concave at some u, then any $x \in \mathcal{E}^u$ does not lie in $\mathcal{E}_\beta$ for any β (nonequivalence of ensembles). This verbal description is made precise in Theorems 4.4 and 4.6, while Theorems

4.4 and 4.8 give necessary and sufficient conditions for equivalence of ensembles to hold. The content of Theorem 4.6 is summarized in Figure 1(a). The contents of Theorems 4.4 and 4.8 are summarized in Figure 1(b).

Theorem 4.4 gives a geometric condition that is necessary and sufficient for equivalence of ensembles to hold. We define C to be the set of $u \in I\!R^\sigma$ for which there exists a supporting hyperplane to the graph of s at $(u, s(u))$. In symbols,

$$C \doteq \{u \in I\!R^\sigma : \exists \beta \in I\!R^\sigma \ni s(w) \leq s(u) + \langle \beta, w - u \rangle \text{ for all } w \in I\!R^\sigma\}. \tag{4.7}$$

If $u \in C$, then the β appearing in this display is a normal vector to the supporting hyperplane. According to part (a) of Theorem 4.4, for a particular $u \in \operatorname{dom} s$ equivalence of ensembles holds if and only if $u \in C$. According to part (b) of the theorem, for a particular $u \in \operatorname{dom} s$ nonequivalence of ensembles holds if and only if $u \notin C$.

Theorem 4.8 refines part (a) of Theorem 4.4 by giving a geometric condition that is necessary and sufficient for full equivalence of ensembles to hold. We define T to be the set of $u \in I\!R^\sigma$ for which there exists a supporting hyperplane to the graph of s that touches the graph of s only at $(u, s(u))$. In symbols,

$$T \doteq \{u \in I\!R^\sigma : \exists \beta \in I\!R^\sigma \ni s(w) < s(u) + \langle \beta, w - u \rangle \text{ for all } w \neq u\}. \tag{4.8}$$

Clearly, T is a subset of C, which is the set of u for which equivalence of ensembles holds [Thm. 4.4(a)]. According to Theorem 4.8, for a particular $u \in \operatorname{dom} s$ full equivalence of ensembles holds if and only if $u \in T$.

Before proving any results on the equivalence and nonequivalence of ensembles, we point out an alternate representation of C that will elucidate the connection between these results and concavity properties of s and s^{**}. In general s is not concave on $I\!R^\sigma$. According to part (b) of Lemma 4.1, C equals the set of $u \in \operatorname{dom} \partial s^{**}$ at which s is concave; i.e., the set of $u \in \operatorname{dom} \partial s^{**}$ such that $s(u)$ equals the value at u of the concave function s^{**}. It follows from part (b) of Lemma 4.1 that if s is not concave at some $u \in \operatorname{dom} s$, then $u \notin C$ and so nonequivalence of ensembles holds [Thm. 4.4 (b)].

It is easy to find a sufficient condition on s^{**} for full equivalence of ensembles to hold. Suppose that for some $u \in I\!R^\sigma$ $s(u) = s^{**}(u)$ and that there exists $\beta \in I\!R^\sigma$ such that

$$s^{**}(w) < s^{**}(u) + \langle \beta, w - u \rangle \text{ for all } w \neq u; \tag{4.9}$$

i.e., the inequality (4.4) defining $\beta \in \partial s^{**}(u)$ holds with strict inequality for all $w \neq u$. Since $s(w) \leq s^{**}(w)$, it follows that

$$s(w) < s(u) + \langle \beta, w - u \rangle \text{ for all } w \neq u. \tag{4.10}$$

That is, u lies in T, which according to Theorem 4.8 is the subset of $I\!R^\sigma$ for which full equivalence of ensembles holds. If, for example, s^{**} is strictly concave in a neighborhood of u, then (4.9) holds for any $\beta \in \partial s^{**}(u)$ and thus we have full equivalence of ensembles.

In order to find a sufficient condition on s^{**} for partial equivalence of ensembles to hold, let u be a point in $I\!R^\sigma$ such that s^{**} is affine in a neighborhood of u. Then except in pathological cases, for any $\beta \in I\!R^\sigma$ the strict inequality (4.10) cannot be valid for all $w \neq u$, and so partial equivalence of ensembles holds.

Part (b) of the next lemma gives the alternate representation of C to which we referred three paragraphs earlier. This representation involves the set

$$\Gamma \doteq \{u \in I\!R^\sigma : s(u) = s^{**}(u)\}.$$

Lemma 4.1. (a) *For u and β in $I\!R^\sigma$, $s(w) \leq s(u) + \langle \beta, w-u \rangle$ for all $w \in I\!R^\sigma$ if and only if both $s(u) = s^{**}(u)$ and $\beta \in \partial s^{**}(u)$.*

(b) *$C = \Gamma \cap \operatorname{dom} \partial s^{**}$, and $C \subset \Gamma \cap \operatorname{dom} s$.*

Remark 4.2. It is not difficult to refine the second assertion in part (b) of this lemma by showing that

$$\Gamma \cap \operatorname{ri}(\operatorname{dom} s) \subset C = \Gamma \cap \operatorname{dom} \partial s^{**} \subset \Gamma \cap \operatorname{dom} s.$$

This relationship implies that, except possibly for relative boundary points of $\operatorname{dom} s$, C consists of $u \in \operatorname{dom} s$ for which $s(u) = s^{**}(u)$. According to Theorem 4.4, equivalence of ensembles holds for a particular $u \in \operatorname{dom} s$ if and only if $u \in C$. Combining this with the observation in the preceding sentence, we see that, except possibly for relative boundary points of $\operatorname{dom} s$, equivalence of ensembles holds for $u \in \operatorname{dom} s$ if and only if $s(u) = s^{**}(u)$.

Proof of Lemma 4.1. (a) We start the proof by first assuming that $s(w) \leq s(u) + \langle \beta, w-u \rangle$ for all $w \in I\!R^\sigma$. It follows that $u \in \operatorname{dom} s$ and that $\langle \beta, u \rangle - s(u) \leq \langle \beta, w \rangle - s(w)$ for all $w \in I\!R^\sigma$. Therefore

$$\langle \beta, u \rangle - s(u) = \inf_{w \in I\!R^\sigma} \{\langle \beta, w \rangle - s(w)\} = \varphi(\beta).$$

Since $s^{**}(w) = \inf_{\gamma \in I\!R^\sigma} \{\langle \gamma, w \rangle - \varphi(\gamma)\} \leq \langle \beta, w \rangle - \varphi(\beta)$, the last display and the inequality $s(u) \leq s^{**}(u)$ imply that for all $w \in I\!R^\sigma$

$$\begin{aligned} s^{**}(w) &\leq \langle \beta, w \rangle - \varphi(\beta) = s(u) + \langle \beta, w \rangle - \langle \beta, u \rangle \\ &\leq s^{**}(u) + \langle \beta, w-u \rangle. \end{aligned}$$

Thus $\beta \in \partial s^{**}(u)$. Setting $w = u$ yields $s(u) = s^{**}(u)$.

Now assume that $s(u) = s^{**}(u)$ and that $\beta \in \partial s^{**}(u)$; thus for all $w \in I\!R^\sigma$

$$s^{**}(w) \leq s^{**}(u) + \langle \beta, w-u \rangle = s(u) + \langle \beta, w-u \rangle.$$

Since $s(w) \leq s^{**}(w)$ for all $w \in I\!R^\sigma$, it follows that for all $w \in I\!R^\sigma$

$$s(w) \leq s(u) + \langle \beta, w-u \rangle.$$

This completes the proof of part (a).

(b) The first assertion is an immediate consequence of part (a). As mentioned in the proof of part (a), if $u \in C$, then $u \in \operatorname{dom} s$. We conclude that $C \subset \Gamma \cap \operatorname{dom} s$, as claimed. ■

The next lemma will facilitate the proofs of a number of our results on the equivalence and nonequivalence of ensembles. Part (b) refines one of the conditions in part (a), substituting a weaker hypothesis that leads to the same conclusion.

Lemma 4.3. *For u and $\beta \in I\!R^\sigma$ the following conclusions hold.*

(a) *The inequality $s(w) \leq s(u) + \langle \beta, w - u \rangle$ is valid for all $w \in I\!R^\sigma$ if and only if $\mathcal{E}^u \neq \emptyset$ and $\mathcal{E}^u \subset \mathcal{E}_\beta$.*

(b) *If $\mathcal{E}^u \cap \mathcal{E}_\beta \neq \emptyset$, then $s(w) \leq s(u) + \langle \beta, w - u \rangle$ for all $w \in I\!R^\sigma$.*

Proof. We first prove that if $s(w) \leq s(u) + \langle \beta, w - u \rangle$ for all $w \in I\!R^\sigma$, then $\mathcal{E}^u \neq \emptyset$ and $\mathcal{E}^u \subset \mathcal{E}_\beta$. The hypothesis implies that $u \in \operatorname{dom} s$ and that $\langle \beta, u \rangle - s(u) \leq \langle \beta, w \rangle - s(w)$ for all $w \in I\!R^\sigma$. Therefore

$$\langle \beta, u \rangle - s(u) = \inf_{w \in I\!R^\sigma} \{\langle \beta, w \rangle - s(w)\} = \varphi(\beta) = \inf_{y \in \mathcal{X}} \{\langle \beta, \tilde{H}(y) \rangle + I(y)\}.$$

The fact that u is an element of $\operatorname{dom} s$ implies that $\mathcal{E}^u \neq \emptyset$. Let x be an arbitrary element in $\mathcal{E}^u$. Since $\tilde{H}(x) = u$ and $I(x) = -s(u)$, the display implies that

$$\langle \beta, \tilde{H}(x) \rangle + I(x) = \inf_{y \in \mathcal{X}} \{\langle \beta, \tilde{H}(y) \rangle + I(y)\}$$

and thus that $x \in \mathcal{E}_\beta$. Since x is an arbitrary element in $\mathcal{E}^u$, it follows that $\mathcal{E}^u \subset \mathcal{E}_\beta$.

In order to complete the proof of part (a), it suffices to prove part (b). Thus suppose that $\mathcal{E}^u \cap \mathcal{E}_\beta \neq \emptyset$ and let x be an arbitrary element in $\mathcal{E}^u \cap \mathcal{E}_\beta$. Since $\mathcal{E}^u \neq \emptyset$, we have $u \in \operatorname{dom} s$. In addition, since $\tilde{H}(x) = u$, $I(x) = -s(u)$, and

$$\langle \beta, \tilde{H}(x) \rangle + I(x) = \inf_{y \in \mathcal{X}} \{\langle \beta, \tilde{H}(y) \rangle + I(y)\} = \varphi(\beta),$$

it follows that for all $w \in I\!R^\sigma$

$$\langle \beta, u \rangle - s(u) = \varphi(\beta) = \inf_{w' \in I\!R^\sigma} \{\langle \beta, w' \rangle - s(w')\} \leq \langle \beta, w \rangle - s(w).$$

Therefore $s(w) \leq s(u) + \langle \beta, w - u \rangle$ for all $w \in I\!R^\sigma$, as claimed. ■

The next theorem is our first main result. Part (a) states that for a particular $u \in \operatorname{dom} s$ equivalence of ensembles holds if and only if $u \in C$. In Theorem 4.9 we make explicit the connection between part (a) and the relationship between thermodynamic equivalence of ensembles and equivalence of ensembles at the level of equilibrium macrostates. Part (b) of the next theorem states that for a particular $u \in \operatorname{dom} s$ nonequivalence of ensembles holds if and only if $u \notin C$. In particular, if s is not concave at some $u \in \operatorname{dom} \partial s^{**}$, then the ensembles are nonequivalent at the level of equilibrium macrostates. Theorem 4.4 was inspired by, and greatly improves upon, the presentation on pages 857-859 of [19], which treats the regularized point vortex model. While part (b) of Theorem 4.4 is related to part (b) of Lemma 5.1 in [32], our Theorem 4.4 makes the nonequivalence of ensembles more explicit.

Theorem 4.4. *We assume Hypotheses* 2.1 *and* 2.2*. For $u \in \operatorname{dom} s$ the following conclusions hold.*

(a) *$u \in C$ if and only if $\mathcal{E}^u \subset \mathcal{E}_\beta$ for some $\beta \in I\!R^\sigma$.*

(b) *$u \notin C$ if and only if $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \in I\!R^\sigma$.*

Proof. (a) This is an immediate consequence of part (a) of Lemma 4.3.

(b) If $u \not\in C$, then for any $\beta \in I\!\!R^\sigma$ the inequality $s(w) \leq s(u) + \langle \beta, w-u \rangle$ does not hold for all $w \in I\!\!R^\sigma$. Part (b) of Lemma 4.3 implies that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \in I\!\!R^\sigma$. To show the converse, assume that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \in I\!\!R^\sigma$ and that $u \in C$. But if $u \in C$, then part (a) of Lemma 4.3 implies that $\mathcal{E}^u \subset \mathcal{E}_\beta$ for some $\beta \in I\!\!R^\sigma$. This contradiction shows that $u \not\in C$, completing the proof. ■

In the next proposition we refine part (a) of Theorem 4.4 by specifying the set of β for which $\mathcal{E}^u \subset \mathcal{E}_\beta$.

Proposition 4.5. *We assume Hypotheses* 2.1 *and* 2.2. *Then for $u \in C$, $\mathcal{E}^u \subset \mathcal{E}_\beta$ for all $\beta \in \partial s^{**}(u)$ and $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \not\in \partial s^{**}(u)$.*

Proof. For $u \in C$, part (b) of Lemma 4.1 implies that $s(u) = s^{**}(u)$ and $\partial s^{**}(u) \neq \emptyset$. If $\beta \in \partial s^{**}(u)$, then part (a) of the same lemma implies that $s(w) \leq s(u) + \langle \beta, w-u \rangle$ for all $w \in I\!\!R^\sigma$. Part (a) of Lemma 4.3 then implies that $\mathcal{E}^u \subset \mathcal{E}_\beta$. This proves the first half of the proposition. On the other hand, if $\beta \not\in \partial s^{**}(u)$, then it is not true that $s(w) \leq s(u) + \langle \beta, w-u \rangle$ for all $w \in I\!\!R^\sigma$ [Lem. 4.1(a)]. It follows from part (b) of Lemma 4.3 that $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$. ■

Theorem 4.4 considers $u \in \mathrm{dom}\, s$, proving that partial or full equivalence of ensembles holds if and only if $u \in C$. The next theorem is our second main result. It shifts focus from $u \in \mathrm{dom}\, s$ to $\beta \in I\!\!R^\sigma$, proving that every set $\mathcal{E}_\beta$ of canonical equilibrium macrostates is a disjoint union of $\mathcal{E}^u$ for u in a particular index set that depends on β.

Theorem 4.6. *We assume Hypotheses* 2.1 *and* 2.2. *Then for all $\beta \in I\!\!R^\sigma$, $\tilde{H}(\mathcal{E}_\beta) \subset \mathrm{dom}\, s$ and*

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

The sets $\mathcal{E}^u$, $u \in \tilde{H}(\mathcal{E}_\beta)$, are nonempty and disjoint.

Proof. Let x be an arbitrary element in $\mathcal{E}_\beta$ and define $\tilde{u} \doteq \tilde{H}(x)$. Since $I_\beta(x) = 0$, we have

$$I(x) + \langle \beta, \tilde{H}(x) \rangle = \inf_{y \in \mathcal{X}} \{ I(y) + \langle \beta, \tilde{H}(y) \rangle \} < \infty,$$

and so $s(\tilde{u}) \geq -I(x) > -\infty$. Thus $\tilde{u} \in \mathrm{dom}\, s$. Because x is an arbitrary element in $\mathcal{E}_\beta$, this proves that $\tilde{H}(\mathcal{E}_\beta) \subset \mathrm{dom}\, s$. Since $\tilde{u} \in \mathrm{dom}\, s$, $\mathcal{E}^{\tilde{u}}$ can be characterized as the set of $x \in \mathcal{X}$ satisfying $\tilde{H}(x) = \tilde{u}$ and $I(x) = -s(\tilde{u})$.

We now prove that $x \in \mathcal{E}^{\tilde{u}}$. Since $x \in \mathcal{E}_\beta$, it follows that for any $y \in \mathcal{X}$

$$I(x) + \langle \beta, \tilde{u} \rangle = I(x) + \langle \beta, \tilde{H}(x) \rangle \leq I(y) + \langle \beta, \tilde{H}(y) \rangle,$$

and thus for any $y \in \mathcal{X}$ satisfying $\tilde{H}(y) = \tilde{u}$, we have $I(x) \leq I(y)$. This implies that

$$I(x) \leq \inf\{I(y) : y \in \mathcal{X}, \tilde{H}(y) = \tilde{u}\} = -s(\tilde{u}) \leq I(x),$$

and so $I(x) = -s(\tilde{u})$. It follows that $x \in \mathcal{E}^{\tilde{u}}$. Since x is an arbitrary element in $\mathcal{E}_\beta$, we have shown that

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

In order to prove the reverse inclusion, we show that for any $u \in \tilde{H}(\mathcal{E}_\beta)$ we have $\mathcal{E}^u \subset \mathcal{E}_\beta$. Any such u has the form $u = \tilde{H}(y)$ for some $y \in \mathcal{E}_\beta$. From our work in the preceding two paragraphs we know that $u \in \operatorname{dom} s$ and $y \in \mathcal{E}^u$. Thus $y \in \mathcal{E}^u \cap \mathcal{E}_\beta$. Since $\mathcal{E}^u \cap \mathcal{E}_\beta \neq \emptyset$, it follows from Theorem 4.4 that $\mathcal{E}^u \subset \mathcal{E}_\beta$. This completes the proof of the display in the theorem.

The sets $\mathcal{E}^u, u \in \tilde{H}(\mathcal{E}_\beta)$, are nonempty since any such u lies in $\operatorname{dom} s$. The sets are also disjoint since for $u \neq u', x \in \mathcal{E}^u \cap \mathcal{E}^{u'}$ implies that $\tilde{H}(x)$ equals both u and u'. The proof of the theorem is complete. ■

The following useful corollary states that when $\mathcal{E}_\beta$ consists of a unique point x, then with $\tilde{u} \doteq \tilde{H}(x)$, $\mathcal{E}^{\tilde{u}}$ consists of the unique point x. This follows from Theorem 4.6 since $\tilde{H}(\mathcal{E}_\beta) = \{\tilde{H}(x)\}$. The corollary sharpens the result on page 861 of [19], which needs the additional hypotheses that s is strictly concave and essentially smooth in order to reach the same conclusion.

Corollary 4.7. *Suppose that $\mathcal{E}_\beta = \{x\}$ for some $\beta \in I\!\!R^\sigma$. Then $\mathcal{E}^{\tilde{u}} = \{x\}$, where $\tilde{u} \doteq \tilde{H}(x)$.*

We now turn our attention to a criterion for full equivalence of ensembles, which is stated in terms of the set T defined in (4.8). Part (a) of Theorem 4.4 states that for a particular $u \in \operatorname{dom} s$ equivalence of ensembles holds if and only if $u \in C$. The next theorem refines this by showing that full equivalence of ensembles holds if and only if $u \in T$. Part (a) gives the sufficiency and part (b) the necessity.

Theorem 4.8. *We assume Hypotheses* 2.1 *and* 2.2. *The following conclusions hold.*

(a) *If $u \in T$, then there exists $\beta \in \partial s^{**}(u)$ such that $\mathcal{E}^u = \mathcal{E}_\beta$.*

(b) *If $u \in C \setminus T$, then $\mathcal{E}^u \subsetneq \mathcal{E}_\beta$ for all $\beta \in \partial s^{**}(u)$ and $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \notin \partial s^{**}(u)$.*

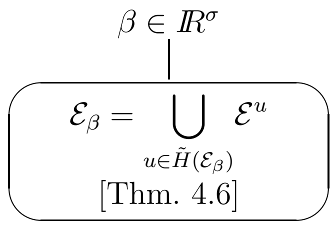
Proof. (a) If $u \in T$, then there exists $\beta \in I\!\!R^\sigma$ such that $s(w) < s(u) + \langle \beta, w - u \rangle$ for all $w \neq u$. Part (a) of Lemma 4.3 implies that $\mathcal{E}^u \subset \mathcal{E}_\beta$. Suppose that $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$. Then Theorem 4.6 implies the existence of $u' \neq u$ such that $\mathcal{E}^{u'} \neq \emptyset$ and $\mathcal{E}^{u'} \subset \mathcal{E}_\beta$, and part (a) of Lemma 4.3 yields

$$s(w) \leq s(u') + \langle \beta, w - u' \rangle \text{ for all } w \in I\!\!R^\sigma.$$

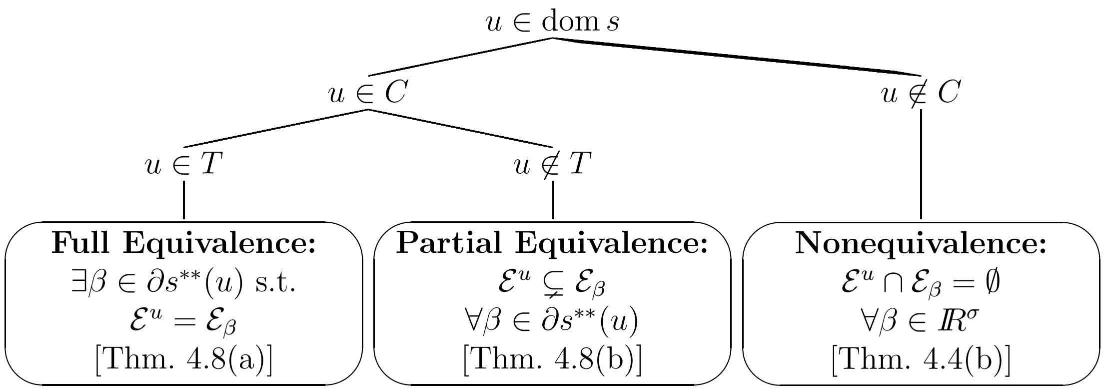
Setting $w = u$ and using the fact that $s(u') < s(u) + \langle \beta, u' - u \rangle$, we see that

$$s(u) \leq s(u') + \langle \beta, u - u' \rangle < s(u) + \langle \beta, u' - u \rangle + \langle \beta, u - u' \rangle = s(u).$$

This contradiction shows that the assumption that $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$ is false. The proof of part (a) is complete.



(a) For $\beta \in I\!R^\sigma$, any $x \in \mathcal{E}_\beta$ lies in some $\mathcal{E}^u$.



(b) There are three possibilities for $u \in \operatorname{dom} s$. The two branches on the left lead to equivalence results, whereas the other branch leads to a nonequivalence result. The sets C and T are defined in (4.7) and (4.8).

Figure 1: Equivalence and nonequivalence of ensembles.

(b) For $u \in C \setminus T$, Proposition 4.5 implies that $\mathcal{E}^u \subset \mathcal{E}_\beta$ for all $\beta \in \partial s^{**}(u)$ and $\mathcal{E}^u \cap \mathcal{E}_\beta = \emptyset$ for all $\beta \notin \partial s^{**}(u)$. We now show that for any $\beta \in \partial s^{**}(u)$, $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$. Since $\mathcal{E}^u \subset \mathcal{E}_\beta$, part (a) of Lemma 4.3 implies that $s(w) \leq s(u) + \langle \beta, w - u \rangle$ for all $w \in I\!R^\sigma$. Since $u \notin T$, there exists $u' \neq u$ such that $s(u') = s(u) + \langle \beta, u' - u \rangle$. Then for all $w \in I\!R^\sigma$

$$s(w) \leq s(u) + \langle \beta, w - u \rangle = s(u') + \langle \beta, w - u' \rangle.$$

It now follows from part (a) of Lemma 4.3 that $\mathcal{E}^{u'} \neq \emptyset$ and $\mathcal{E}^{u'} \subset \mathcal{E}_\beta$. Thus $\mathcal{E}^u$ is a proper subset of $\mathcal{E}_\beta$, as claimed. ■

We recall that thermodynamic equivalence of ensembles is said to hold when s is concave on $I\!R^\sigma$. The next theorem addresses the issue of how thermodynamic equivalence of ensembles mirrors equivalence of ensembles at the level of equilibrium macrostates. Part (a) shows that thermodynamic equivalence is a sufficient condition for macroscopic equivalence to hold for all $u \in \operatorname{dom} \partial s$. Since when s is concave on $I\!R^\sigma$ we have $\operatorname{ri}(\operatorname{dom} s) \subset \operatorname{dom} \partial s \subset \operatorname{dom} s$, it follows that thermodynamic equivalence is a sufficient condition for macroscopic equivalence to hold for all $u \in \operatorname{dom} s$ except possibly for relative boundary points. Part (b) proves a partial converse to (a). In part (c) we point out that thermodynamic equivalence is equivalent to macroscopic equivalence under an extra hypothesis on the domains of s, s^{**}, and ∂s^{**}. The proof of the theorem follows readily from our previous results. The theorem is related to Lemma 6.2 and Theorem 6.1 in [32].

Theorem 4.9. (a) *Assume that s is concave on $I\!\!R^\sigma$. Then for all $u \in \operatorname{dom} \partial s$, $\mathcal{E}^u \subset \mathcal{E}_\beta$ for some $\beta \in \partial s(u)$. Thus, thermodynamic equivalence of ensembles implies equivalence of ensembles at the level of equilibrium macrostates for all $u \in \operatorname{dom} \partial s$.*

(b) *Assume that $\operatorname{dom} s = \operatorname{dom} s^{**}$ and that for all $u \in \operatorname{dom} s$ there exists $\beta \in I\!\!R^\sigma$ such that $\mathcal{E}^u \subset \mathcal{E}_\beta$. Then s is concave on $I\!\!R^\sigma$. Thus, under the hypothesis that $\operatorname{dom} s = \operatorname{dom} s^{**}$, equivalence of ensembles at the level of equilibrium macrostates for all $u \in \operatorname{dom} s$ implies thermodynamic equivalence of ensembles.*

(c) *Assume that $\operatorname{dom} s = \operatorname{dom} s^{**} = \operatorname{dom} \partial s^{**}$. Then thermodynamic equivalence of ensembles holds if and only if the ensembles are equivalent at the level of equilibrium macrostates.*

Proof. (a) If s is concave on $I\!\!R^\sigma$, then $s = s^{**}$ on $I\!\!R^\sigma$ and $C = \operatorname{dom} \partial s^{**} = \operatorname{dom} \partial s$ [Lem. 4.1(b)]. Part (a) of Theorem 4.3 completes the proof of part (a).

(b) The hypotheses imply that any element of $\operatorname{dom} s$ is an element of C, which in turn is a subset of $\Gamma \doteq \{u \in I\!\!R^\sigma : s(u) = s^{**}(u)\}$. It follows that s and s^{**} agree on $\operatorname{dom} s = \operatorname{dom} s^{**}$ and thus that s is concave on $I\!\!R^\sigma$.

(c) This follows from parts (a) and (b). ■

With Theorem 4.9 the presentation of the main results in this section is complete. We end this section by giving two additional theorems in which we explore further relationships involving $\mathcal{E}_\beta$, $\mathcal{E}^u$, and the thermodynamic functions φ and s.

In part (a) of the next theorem we refine Theorem 4.6 by proving that $\mathcal{E}_\beta = \bigcup_{u \in \partial\varphi(\beta) \cap \Gamma} \mathcal{E}^u$, where $\partial\varphi(\beta)$ denotes the superdifferential at β of the concave function φ and, as introduced in Lemma 4.1, $\Gamma \doteq \{u \in I\!\!R^\sigma : s(u) = s^{**}(u)\}$. This in turn allows us to give, in part (b), a necessary and sufficient condition for the differentiability of φ at a point β. Part (c) is a special case of part (b).

Theorem 4.10. *We assume Hypotheses 2.1 and 2.2. The following conclusions hold.*

(a) *For all $\beta \in I\!\!R^\sigma$*

$$\mathcal{E}_\beta = \bigcup_{u \in \tilde{H}(\mathcal{E}_\beta)} \mathcal{E}^u = \bigcup_{u \in \partial\varphi(\beta) \cap \Gamma} \mathcal{E}^u.$$

(b) *φ is differentiable at β if and only if both $\mathcal{E}_\beta = \mathcal{E}^u$ for some u and $\partial\varphi(\beta) \subset \Gamma$.*

(c) *If s is concave on $I\!\!R^\sigma$, then φ is differentiable at β if and only if $\mathcal{E}_\beta = \mathcal{E}^u$ for some u.*

Proof. (a) It follows from part (a) of Lemma 4.3 and part (a) of Lemma 4.1 that

$$\mathcal{E}^u \neq \emptyset \text{ and } \mathcal{E}^u \subset \mathcal{E}_\beta \text{ if and only if } s(u) = s^{**}(u) \text{ and } \beta \in \partial s^{**}(u).$$

Since $\beta \in \partial s^{**}(u)$ if and only if $u \in \partial s^*(\beta) = \partial\varphi(\beta)$ [45, p. 218], it follows that

$$\mathcal{E}^u \neq \emptyset \text{ and } \mathcal{E}^u \subset \mathcal{E}_\beta \text{ if and only if } u \in \partial\varphi(\beta) \cap \Gamma. \tag{4.11}$$

Thus

$$\bigcup_{u \in \partial\varphi(\beta)\cap\Gamma} \mathcal{E}^u \subset \mathcal{E}_\beta.$$

We complete the proof of part (a) by showing that we have equality in this display. By Theorem 4.6 $\mathcal{E}_\beta$ is a disjoint union of $\mathcal{E}^u$ for $u \in \tilde{H}(\mathcal{E}_\beta) \subset \operatorname{dom} s$. Hence for each $u \in \tilde{H}(\mathcal{E}_\beta)$, $\mathcal{E}^u \neq \emptyset$ and $\mathcal{E}^u \subset \mathcal{E}_\beta$. Thus (4.11) implies that $\tilde{H}(\mathcal{E}_\beta) \subset \partial\varphi(\beta) \cap \Gamma$. We conclude that

$$\bigcup_{u \in \partial\varphi(\beta)\cap\Gamma} \mathcal{E}^u \subset \mathcal{E}_\beta = \bigcup_{u \in \tilde{H}(\mathcal{E}_\beta)} \mathcal{E}^u \subset \bigcup_{u \in \partial\varphi(\beta)\cap\Gamma} \mathcal{E}^u,$$

and therefore $\bigcup_{u \in \partial\varphi(\beta)\cap\Gamma} \mathcal{E}^u = \mathcal{E}_\beta$.

(b) We first assume that φ is differentiable at β. Since by part (a) $\partial\varphi(\beta) \cap \Gamma \neq \emptyset$ for any β, the differentiability of φ at β implies that $\partial\varphi(\beta) = \{\nabla\varphi(\beta)\} \subset \Gamma$ and that $\mathcal{E}_\beta = \mathcal{E}^{\nabla\varphi(\beta)}$. We now assume that $\mathcal{E}_\beta = \mathcal{E}^u$ for some u and $\partial\varphi(\beta) \subset \Gamma$. Since part (a) implies that $\partial\varphi(\beta) \cap \Gamma = \{u\}$, we conclude that $\partial\varphi(\beta) = \partial\varphi(\beta) \cap \Gamma = \{u\}$ and therefore that φ is differentiable at β.

(c) This follows from part (b) since the concavity of s on $I\!\!R^\sigma$ implies that $\Gamma = I\!\!R^\sigma$, and so $\partial\varphi(\beta) \subset \Gamma$ is always true. ■

The next theorem is the final result in this section. Under the hypothesis that s is concave on $I\!\!R^\sigma$, part (a) gives a simpler form of the representation in part (a) of Theorem 4.10. Part (b) is a partial converse of part (a).

Theorem 4.11. *We assume Hypotheses* 2.1 *and* 2.2. *The following conclusions hold.*

(a) *Assume that s is concave on $I\!\!R^\sigma$. Then for all $\beta \in I\!\!R^\sigma$*

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

(b) *Now assume that for all $\beta \in I\!\!R^\sigma$*

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

Then s is a finite concave function on any convex subset of $\operatorname{ri}(\operatorname{dom} s)$.

Proof. (a) Since s is concave on $I\!\!R^\sigma$, Γ equals $I\!\!R^\sigma$ and thus $\partial\varphi(\beta) \cap \Gamma = \partial\varphi(\beta)$ for all $\beta \in I\!\!R^\sigma$. Hence part (a) follows from part (a) of Theorem 4.10.

(b) Since by definition $\mathcal{E}^u = \emptyset$ for all $u \notin \operatorname{dom} s$, it follows from the hypothesis in part (b) and from part (a) of Theorem 4.10 that for all $\beta \in I\!\!R^\sigma$

$$\mathcal{E}_\beta = \bigcup_{u \in \partial\varphi(\beta)\cap\operatorname{dom} s} \mathcal{E}^u = \bigcup_{u \in \partial\varphi(\beta)\cap\Gamma} \mathcal{E}^u.$$

Thus $\partial\varphi(\beta) \cap \operatorname{dom} s = \partial\varphi(\beta) \cap \Gamma$. Taking the union over all $\beta \in I\!\!R^\sigma$ yields

$$\bigcup_{\beta \in I\!\!R^\sigma} \partial\varphi(\beta) \cap \operatorname{dom} s = \bigcup_{\beta \in I\!\!R^\sigma} \partial\varphi(\beta) \cap \Gamma \subset \Gamma.$$

By standard duality theory for upper semicontinuous, concave functions on $I\!\!R^\sigma$ [45, p. 218], $\bigcup_{\beta \in I\!\!R^\sigma} \partial\varphi(\beta) = \operatorname{dom} \partial s^{**}$. Thus

$$(\operatorname{dom} \partial s^{**}) \cap (\operatorname{dom} s) \subset \Gamma.$$

Since $\operatorname{ri}(\operatorname{dom} s) \subset \operatorname{ri}(\operatorname{dom} s^{**}) \subset \operatorname{dom} \partial s^{**}$, we conclude that $\operatorname{ri}(\operatorname{dom} s) \subset \Gamma$ and therefore that s is concave on any convex subset of $\operatorname{ri}(\operatorname{dom} s)$. The proof of the theorem is complete. ■

In the next section we extend the large deviation theorems in Sections 2 and 3 and the duality theorems in the present section to the study of mixed ensembles.

5 Mixed Ensembles

In broad terms the canonical ensemble differs from the microcanonical ensemble by the manner in which the dynamical invariants are incorporated in the respective probability measures: exponentiation in the former ensemble and conditioning in the latter ensemble. In Section 5.1 we define two classes of mixed ensembles, a mixed canonical-microcanonical ensemble and a mixed microcanonical-canonical ensemble, which differ only in the order in which the exponentiation and the conditioning are performed. In part (b) of Theorem 5.1.1 we show that with respect to both of these ensembles the hidden process Y_n satisfies the large deviation principle with the same rate function. Hence the sets of equilibrium macrostates for both of these ensembles are the same. In Section 5.2 we present complete equivalence and nonequivalence results relating the sets of equilibrium macrostates for the mixed and the pure canonical ensembles. In Section 5.3, we do the same for the sets of equilibrium macrostates for the mixed and the pure microcanonical ensembles. These results will be applied in future work to a number of problems, including soliton turbulence for the nonlinear Schrödinger equation [17].

5.1 Properties of the Mixed Ensembles

The definitions of the mixed ensembles involve quantities introduced in Hypotheses 2.1 and 2.2. We shall use the notation $\mathrm{Can}(H_n; P_n)_\beta$ to denote the canonical ensemble $P_{n,\beta}$, which is defined in (2.1), and the notation $\mathrm{Micro}(H_n; P_n)^{u,r}$ to denote the microcanonical ensemble $P_n^{u,r}$, which is defined in (3.4). The LDP's for Y_n with respect to the canonical ensemble and with respect to the microcanonical ensemble are given in Theorems 2.4 and 3.2, respectively. The respective rate functions are

$$I_\beta(x) \doteq I(x) + \langle \beta, \tilde{H}(x) \rangle - \inf_{y \in \mathcal{X}} \{ I(y) + \langle \beta, \tilde{H}(y) \rangle \}$$

and for $u \in \operatorname{dom} J$

$$I^u(x) \doteq \begin{cases} I(x) - J(u) & \text{if } \tilde{H}(x) = u, \\ \infty & \text{otherwise.} \end{cases}$$

In the sequel we shall use the following alternate formula for I^u:

$$I^u(x) = I(\{x\} \cap \tilde{H}^{-1}(\{u\})) - J(u).$$

Analogous formulas will arise in the study of the mixed ensembles.

In order to introduce the mixed ensembles, we assume that $\sigma \geq 2$. Let τ be an integer satisfying $1 \leq \tau \leq \sigma$ and consider decompositions of H_n and of $\tilde{H}$ defined as follows:

$$H_n = (H_n^1, H_n^2), \text{ where } H_n^1 \doteq (H_{n,1}, \ldots, H_{n,\tau}) \text{ and } H_n^2 \doteq (H_{n,\tau+1}, \ldots, H_{n,\sigma}),$$

$$\tilde{H} = (\tilde{H}^1, \tilde{H}^2), \text{ where } \tilde{H}^1 \doteq (\tilde{H}_1, \ldots, \tilde{H}_\tau) \text{ and } \tilde{H}^2 \doteq (\tilde{H}_{\tau+1}, \ldots, \tilde{H}_\sigma).$$

Writing $\beta = (\beta^1, \beta^2) \in I\!R^\tau \times I\!R^{\sigma-\tau}$ and $u = (u^1, u^2) \in I\!R^\tau \times I\!R^{\sigma-\tau}$, we define

$$\begin{aligned} \mathrm{Can}(H_n^1, H_n^2; P_n)_{\beta^1,\beta^2}(d\omega) &\doteq \mathrm{Can}(H_n; P_n)_\beta(d\omega) \\ &= \frac{1}{Z_n(\beta^1, \beta^2)} \exp[-\langle \beta^1, H_n^1(\omega) \rangle - \langle \beta^2, H_n^2(\omega) \rangle] \, P_n(d\omega), \end{aligned}$$

where $Z_n(\beta^1, \beta^2) \doteq Z_n(\beta)$, and we define

$$\begin{aligned} \mathrm{Micro}(H_n^1, H_n^2; P_n)^{u^1,u^2,r}(d\omega) &\doteq \mathrm{Micro}(H_n; P_n)^{u,r}(d\omega) \\ &= P_n(d\omega | H_n^1 \in \{u^1\}^{(r)}, H_n^2 \in \{u^2\}^{(r)}). \end{aligned}$$

The function $J(u) \doteq \inf\{I(x) : x \in \mathcal{X}, \tilde{H}(x) = u\}$ plays a key role in the large deviation analysis of the microcanonical ensemble. We rewrite this function as

$$J(u^1, u^2) \doteq \inf\{I(x) : x \in \mathcal{X}, \tilde{H}^1(x) = u^1, \tilde{H}^2(x) = u^2\}. \tag{5.1.1}$$

The innovation of the present subsection is to consider the asymptotic properties of two mixed ensembles, both at the level of thermodynamic functions and at the level of equilibrium macrostates. We define a mixed canonical-microcanonical ensemble by replacing the measure P_n in the canonical ensemble $\mathrm{Can}(H_n^1; P_n)_{\beta^1}$ by the microcanonical ensemble $\mathrm{Micro}(H_n^2; P_n)^{u^2,r}$. For $u^2 \in I\!\!R^{\sigma-\tau}$ and $\beta^1 \in I\!\!R^{\tau}$, the resulting measure is given by

$$\begin{aligned} &\mathrm{Can}(H_n^1; \mathrm{Micro}(H_n^2; P_n)^{u^2,r})_{\beta^1}(d\omega) \\ &\quad \doteq \frac{1}{Z_n(\beta^1, \{u^2\}^{(r)})} \exp[-\langle \beta^1, H_n^1(\omega)\rangle]\, P_n(d\omega | H_n^2 \in \{u^2\}^{(r)}), \end{aligned}$$

where

$$Z_n(\beta^1, \{u^2\}^{(r)}) \doteq \int_{\Omega_n} \exp[-\langle \beta^1, H_n^1(\omega)\rangle]\, P_n(d\omega | H_n^2 \in \{u^2\}^{(r)}).$$

By a similar verification as in the paragraph after Proposition 3.1, the microcanonical ensemble $\mathrm{Micro}(H_n^2; P_n)^{u^2,r}$, and thus this mixed ensemble, are well defined for all sufficiently large n provided u^2 lies in the domain of

$$J^2(u^2) \doteq \inf\{I(x) : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\}. \tag{5.1.2}$$

In an analogous way, we define a mixed microcanonical-canonical ensemble by replacing the measure P_n in the microcanonical ensemble $\mathrm{Micro}(H_n^2; P_n)^{u^2,r}$ by the canonical ensemble $\mathrm{Can}(H_n; P_n)_{\beta^1}$. For $\beta^1 \in I\!\!R^{\tau}$ and $u^2 \in I\!\!R^{\sigma-\tau}$, the resulting measure is given by

$$\mathrm{Micro}(H_n^2; \mathrm{Can}(H_n^1; P_n)_{\beta^1})^{u^2,r}(d\omega) \doteq Q_{n,\beta^1}(d\omega | H_n^2 \in \{u^2\}^{(r)}),$$

where

$$Q_{n,\beta^1}(d\omega) \doteq \frac{1}{Z_n(\beta^1)} \exp[-\langle \beta^1, H_n^1(\omega)\rangle]\, P_n(d\omega).$$

This mixed ensemble is well defined for all sufficiently large n provided u^2 lies in the domain of the function J_{β^1} that stands in the same relationship to the mixed ensemble as the function J in (5.1.1) stands to the microcanonical ensemble. Since J is defined in terms of I, which is the rate function in the LDP for Y_n with respect to P_n, J_{β^1} is defined in terms of the rate function for Y_n with respect to the canonical ensemble $\mathrm{Can}(H_n^1; P_n)_{a_n\beta^1}$. By Theorem 2.4, this rate function is given by

$$I_{\beta^1}(x) \doteq I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle - \inf_{y \in \mathcal{X}}\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle\}.$$

It follows that

$$\begin{aligned} J_{\beta^1}(u^2) &\doteq \inf\{I_{\beta^1}(x) : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\} \\ &= \inf\{I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\} \\ &\qquad - \inf_{y\in\mathcal{X}}\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle\}. \end{aligned} \tag{5.1.3}$$

By the discussion earlier in this paragraph, the mixed ensemble $\mathrm{Micro}(H_n^2; \mathrm{Can}(H_n^1; P_n)_{\beta^1})^{u^2,r}$ is well-defined for all sufficiently large n provided u^2 lies in the domain of J_{β^1}. Since $\tilde{H}^1(x)$ is finite for all $x \in \mathcal{X}$, $u^2 \in \mathrm{dom}\, J_{\beta^1}$ if and only if $u^2 \in \mathrm{dom}\, J^2$. By the same proof as that of Proposition 3.1, with respect to P_n, the sequences $\tilde{H}^2(Y_n)$ and H_n^2 satisfy the LDP on $I\!\!R^{\sigma-\tau}$ with rate function J^2. As a consequence, $\mathrm{dom}\, J^2$ is nonempty as is $\mathrm{dom}\, J_{\beta^1}$.

We recall from Section 4 that

$$s(u) \doteq -J(u) = -\inf\{I(x) : x \in \mathcal{X}, \tilde{H}(x) = u\}$$

defines the microcanonical entropy and that its Legendre-Fenchel transform gives the canonical free energy. Both functions appear in relationships involving $\mathcal{E}_\beta$ and $\mathcal{E}^u$ that appear in that section. In an analogous way, for $\beta^1 \in I\!\!R^\tau$ and $u^2 \in I\!\!R^{\sigma-\tau}$, we define the entropy with respect to the mixed ensemble $\mathrm{Micro}(H_n^2; \mathrm{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$ to be

$$s_{\beta^1}(u^2) \doteq -J_{\beta^1}(u^2). \tag{5.1.4}$$

This entropy and the associated free energy will appear in the results on equivalence and nonequivalence of ensembles to be given in Section 5.2.

In order to complete the definitions of the various ensembles, we also consider the pure ensembles

$$\mathrm{Can}(H_n^1; \mathrm{Can}(H_n^2; P_n)_{\beta^2})_{\beta^1} \quad \text{and} \quad \mathrm{Micro}(H_n^1; \mathrm{Micro}(H_n^2; P_n)^{u^2,r})^{u^1,r},$$

which are defined similarly as above. We omit the simple calculation showing that for all n and r

$$\mathrm{Can}(H_n^1; \mathrm{Can}(H_n^2; P_n)_{\beta^2})_{\beta^1}(d\omega) = \mathrm{Can}(H_n^1, H_n^2; P_n)_{\beta^1,\beta^2}(d\omega) \tag{5.1.5}$$

and

$$\mathrm{Micro}(H_n^1; \mathrm{Micro}(H_n^2; P_n)^{u^2,r})^{u^1,r}(d\omega) = \mathrm{Micro}(H_n^1, H_n^2; P_n)^{u^1,u^2,r}(d\omega). \tag{5.1.6}$$

On the other hand, for all n and r the mixed canonical-microcanonical ensemble and the mixed microcanonical-canonical ensemble are different. In the next theorem we record the LDP's satisfied by Y_n with respect to the various ensembles introduced in this subsection. The pleasant surprise is that although the two mixed ensembles are different for all n and r, with respect to each of them, with β^1 replaced by $a_n\beta^1$, Y_n satisfies the LDP with the identical rate function.

Before stating the theorem, we define the rate functions for each ensemble. For $\beta = (\beta^1, \beta^2) \in I\!\!R^\tau \times I\!\!R^{\sigma-\tau}$, $u^2 \in \mathrm{dom}\, J^2$, and $u = (u^1, u^2) \in \mathrm{dom}\, J$, we define the following functions mapping $\mathcal{X}$ into $[0,\infty]$:

$$\begin{aligned} I_{\beta^1,\beta^2}(x) &\doteq I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle + \langle \beta^2, \tilde{H}^2(x)\rangle \\ &\qquad - \inf_{y\in\mathcal{X}}\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle + \langle \beta^2, \tilde{H}^2(y)\rangle\}, \end{aligned} \tag{5.1.7}$$

$$
\begin{aligned}
I_{\beta^1}^{u^2}(x) \;&\doteq\; I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) + \langle \beta^1, \tilde{H}^1(x) \rangle \\
&\qquad - \inf\{I(y) + \langle \beta^1, \tilde{H}^1(y) \rangle : y \in \mathcal{X}, \tilde{H}^2(y) = u^2\},
\end{aligned} \tag{5.1.8}
$$

and

$$
I^{u^1,u^2}(x) \doteq I(\{x\} \cap (\tilde{H}^1)^{-1}(\{u^1\}) \cap (\tilde{H}^2)^{-1}(\{u^2\})) - J(u^1, u^2). \tag{5.1.9}
$$

Theorem 5.1.1. *We assume Hypotheses* 2.1 *and* 2.2. *For* $(\beta^1, \beta^2) \in I\!R^\tau \times I\!R^{\sigma-\tau}$ *the following conclusions hold.*

(a) *With respect to the canonical ensemble* $\mathrm{Can}(H_n^1, H_n^2; P_n)_{a_n\beta^1, a_n\beta^2}$, Y_n *satisfies the LDP on* $\mathcal{X}$ *with rate function* I_{β^1,β^2} *given in* (5.1.7).

(b) *Take* $u^2 \in \mathrm{dom}\, J^2$ [*see* (5.1.2)]. *Both with respect to the mixed canonical-microcanonical ensemble* $\mathrm{Can}(H_n^1; \mathrm{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$ *and with respect to the mixed microcanonical-canonical ensemble* $\mathrm{Micro}(H_n^2; \mathrm{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$, Y_n *satisfies the LDP on* $\mathcal{X}$, *in the double limit* $n \to \infty$ *and* $r \to 0$, *with rate function* $I_{\beta^1}^{u^2}$ *given in* (5.1.8).

(c) *Take* $u = (u^1, u^2) \in \mathrm{dom}\, J$ [*see* (5.1.1)]. *With respect to the microcanonical ensemble* $\mathrm{Micro}(H_n^1, H_n^2; P_n)^{u^1,u^2,r}$, Y_n *satisfies the LDP on* $\mathcal{X}$, *in the double limit* $n \to \infty$ *and* $r \to 0$, *with rate function* I^{u^1,u^2} *given in* (5.1.9).

Proof. Part (a) is proved in Theorem 2.4, and part (c) is proved in Theorem 3.2. In part (b) we first prove the LDP for Y_n with respect to $\mathrm{Micro}(H_n^2;\mathrm{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$. Theorem 2.4 implies that with respect to $\mathrm{Can}(H_n^1; P_n)_{a_n\beta^1}$, Y_n satisfies the LDP with rate function

$$
I_{\beta^1}(x) \doteq I(x) + \langle \beta^1, \tilde{H}^1(x) \rangle - \inf_{y \in \mathcal{X}}\{I(y) + \langle \beta^1, \tilde{H}^1(y) \rangle\}.
$$

With P_n replaced by $\mathrm{Can}(H_n^1; P_n)_{a_n\beta^1}$ and I replaced by I_{β^1}, Theorem 3.2 guarantees that if $u^2 \in \mathrm{dom}\, J_{\beta^1} = \mathrm{dom}\, J^2$, then with respect to $\mathrm{Micro}(H_n^2;\mathrm{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$, Y_n satisfies the LDP, in the double limit $n \to \infty$ and $r \to 0$, with rate function

$$
(I_{\beta^1})^{u^2}(x) \doteq I_{\beta^1}(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) - \inf\{I_{\beta^1}(y) : y \in \mathcal{X}, \tilde{H}^2(y) = u^2\}.
$$

Substituting the definition of I_{β^1}, we see that

$$
\begin{aligned}
(I_{\beta^1})^{u^2}(x) \;&=\; I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) + \langle \beta^1, \tilde{H}^1(x) \rangle \\
&\qquad - \inf\{I(y) + \langle \beta^1, \tilde{H}^1(y) \rangle : y \in \mathcal{X}, \tilde{H}^2(y) = u^2\}.
\end{aligned}
$$

This is the function $I_{\beta^1}^{u^2}$ defined in (5.1.8). We have proved that with respect to $\mathrm{Micro}(H_n^2;\mathrm{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$, Y_n satisfies the LDP, in the double limit $n \to \infty$ and $r \to 0$, with rate function $I_{\beta^1}^{u^2}$.

We next consider the LDP for Y_n with respect to $\mathrm{Can}(H_n^1; \mathrm{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$. Since $u^2 \in \mathrm{dom}\, J^2$, Theorem 3.2 implies that with respect to $\mathrm{Micro}(H_n^2; P_n)^{u^2}$, Y_n satisfies the LDP, in the double limit $n \to \infty$ and $r \to 0$, with rate function

$$
I^{u^2}(x) \doteq I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) - J^2(u^2).
$$

One can easily modify the proof of Theorem 2.4 to handle the situation in which P_n is replaced by a doubly indexed class of probability measures such as $\text{Micro}(H_n^2; P_n)^{u^2,r}$ with the property that with respect to these measures Y_n satisfies the LDP. With this modification, replacing P_n by $\text{Micro}(H_n^2; P_n)^{u^2,r}$ and I by I^{u^2}, we see that with respect to $\text{Can}(H_n^1; \text{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$, Y_n satisfies the LDP, in the double limit $n \to \infty$ and $r \to 0$, with rate function

$$\begin{aligned}(I^{u^2})_{\beta^1}(x) &\doteq I^{u^2}(x) + \langle \beta^1, \tilde{H}^1(x)\rangle - \inf_{y \in \mathcal{X}}\{I^{u^2}(y) + \langle \beta^1, \tilde{H}^1(y)\rangle\} \\ &= I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) + \langle \beta^1, \tilde{H}^1(x)\rangle \\ &\quad - \inf\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle : y \in \mathcal{X}, \tilde{H}^2(y) = u^2\}.\end{aligned}$$

This is the function $I_{\beta^1}^{u^2}$ defined in (5.1.8). We have shown that with respect to $\text{Can}(H_n^1; \text{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$, Y_n satisfies the LDP, in the double limit $n \to \infty$ and $r \to 0$, with rate function $I_{\beta^1}^{u^2}$. The proof of the theorem is complete. ■

In the next two subsections, we consider equivalence and nonequivalence results for the ensembles whose LDP's are derived in Theorem 5.1.1. These results are derived as immediate consequences of our work in Section 4, where equivalence and nonequivalence results for the canonical and microcanonical ensembles were derived.

5.2 Equivalence and Nonequivalence of the Canonical and Mixed Ensembles

In this subsection we study, at the level of equilibrium macrostates, the equivalence and nonequivalence of the canonical ensemble $\text{Can}(H_n^1, H_n^2; P_n)_{a_n\beta^1, a_n\beta^2}$ and the mixed ensemble $\text{Micro}(H_n^2; \text{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$. The parameters β^1, β^2, and u^2 satisfy $\beta^1 \in I\!R^\tau$, $\beta^2 \in I\!R^{\sigma-\tau}$, and $u^2 \in \text{dom}\, J^2$, where

$$J^2(u^2) \doteq \inf\{I(x) : \tilde{H}^2(x) = u^2\}.$$

By a similar verification as in the paragraph after Proposition 3.1, this condition on u^2 guarantees that the mixed ensemble is well defined for all sufficiently large n. The relationships between the sets of equilibrium macrostates for the two ensembles follow immediately from Theorems 4.4, 4.6, and 4.8 with minimal changes in proof. Hence we shall only summarize them in Figure 2.

By Theorem 5.1.1, for $(\beta^1, \beta^2) \in I\!R^\tau \times I\!R^{\sigma-\tau}$, with respect to $\text{Can}(H_n^1, H_n^2; P_n)_{a_n\beta^1, a_n\beta^2}$ Y_n satisfies the LDP with rate function

$$\begin{aligned}I_{\beta^1,\beta^2}(x) &\doteq I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle + \langle \beta^2, \tilde{H}^2(x)\rangle \\ &\quad - \inf_{y \in \mathcal{X}}\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle + \langle \beta^2, \tilde{H}^2(y)\rangle\}.\end{aligned} \tag{5.2.1}$$

In addition, for $(\beta^1, u^2) \in I\!R^\tau \times \text{dom}\, J^2$, with respect to $\text{Micro}(H_n^2; \text{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$ Y_n satisfies the LDP with rate function

$$I_{\beta^1}^{u^2}(x) \doteq I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) + \langle \beta^1, \tilde{H}^1(x)\rangle) - \psi_{\beta^1}^{u^2}, \tag{5.2.2}$$

where

$$\psi_{\beta^1}(u^2) \doteq \inf\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle : y \in \mathcal{X}, \tilde{H}^2(y) = u^2\}.$$

For $\beta^1 \in I\!R^\tau$, $\beta^2 \in I\!R^{\sigma-\tau}$, and $u^2 \in \text{dom}\, J^2$, we define the corresponding sets of equilibrium macrostates

$$\mathcal{E}_{\beta^1,\beta^2} \doteq \{x \in \mathcal{X} : I_{\beta^1,\beta^2}(x) = 0\}$$

and

$$\begin{aligned} \mathcal{E}_{\beta^1}^{u^2} &\doteq \{x \in \mathcal{X} : I_{\beta^1}^{u^2}(x) = 0\} \\ &= \{x \in \mathcal{X} : \tilde{H}^2(x) = u^2, I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle = \psi_{\beta^1}^{u^2}\}. \end{aligned}$$

As the sets of points at which the corresponding rate functions attain their minimum of 0, both $\mathcal{E}_{\beta^1,\beta^2}$ and $\mathcal{E}_{\beta^1}^{u^2}$ are nonempty, compact subsets of $\mathcal{X}$ for $\beta^1 \in I\!R^\tau$, $\beta^2 \in I\!R^{\sigma-\tau}$, and $u^2 \in \text{dom}\, J^2$. The main purpose of this subsection is to record the relationships between these sets.

Before doing so, we point out a concentration property, relative to the set $\mathcal{E}_{\beta^1}^{u^2}$, of the distributions of Y_n with respect to the mixed ensemble $\text{Micro}(H_n^2; \text{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$. This concentration property is an immediate consequence of the LDP proved in part (b) of Theorem 5.1.1. It justifies calling $\mathcal{E}_{\beta^1}^{u^2}$ the set of equilibrium macrostates with respect to the mixed ensemble. This concentration property is analogous to those for the canonical ensemble and for the microcanonical ensemble given in part (c) of Theorem 2.4 and in part (b) of Theorem 3.5; the proof is omitted.

Theorem 5.2.1. *We assume Hypotheses* 2.1 *and* 2.2. *For $\beta^1 \in I\!R^\tau$, $u^2 \in \text{dom}\, J^2$, and A any Borel subset of $\mathcal{X}$ whose closure $\bar{A}$ satisfies $\bar{A} \cap \mathcal{E}_{\beta^1}^{u^2} = \emptyset$, we have $I_{\beta^1}^{u^2}(\bar{A}) > 0$. In addition, there exists $r_0 \in (0,1)$ and for all $r \in (0, r_0]$ there exists $C_r < \infty$ such that*

$$\text{Micro}(H_n^2; \text{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}\{Y_n \in A\} \leq C_r \exp[-a_n I_{\beta^1}^{u^2}(\bar{A})/2] \to 0 \quad \textit{as } n \to \infty.$$

As in Theorem 3.6, one can also study compactness and weak limit properties of the distributions of Y_n with respect to $\text{Micro}(H_n^2; \text{Can}(H_n^1; P_n)_{a_n\beta^1})^{u^2,r}$. We shall omit this topic.

We return to the relationships between $\mathcal{E}_{\beta^1,\beta^2}$ and $\mathcal{E}_{\beta^1}^{u^2}$. Since for each n

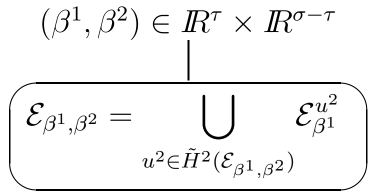
$$\text{Can}(H_n^1, H_n^2; P_n)_{\beta^1,\beta^2} \quad \text{and} \quad \text{Can}(H_n^2; \text{Can}(H_n^1; P_n)_{\beta^1})_{\beta^2}$$

are equal, we can derive the relationships between these sets of equilibrium macrostates by applying the results of Section 4 to the canonical ensemble and microcanonical ensemble

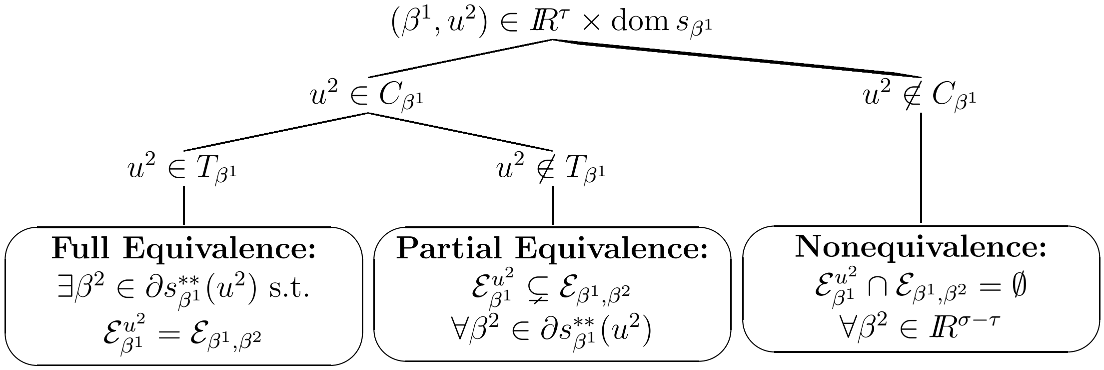
$$\text{Can}(H_n^2; Q_n)_{a_n\beta^2} \quad \text{and} \quad \text{Micro}(H_n^2; Q_n)^{u^2}, \text{ with } Q_n \doteq \text{Can}(H_n^1; P_n)_{a_n\beta^1}.$$

To this end, we introduce the relevant thermodynamic functions. With respect to $\text{Can}(H_n^2; \text{Can}(H_n^1; P_n)_{a_n\beta^1})_{a_n\beta^2}$ the free energy is given by

$$\begin{aligned} \varphi_{\beta^1}(\beta^2) &= -\lim_{n\to\infty} \frac{1}{a_n} \log \int_{\Omega_n} \exp[-a_n\langle \beta^2, H_n^2\rangle]\, d\Big(\text{Can}(H_n^1; P_n)_{a_n\beta^1}\Big) \qquad (5.2.3) \\ &= \inf_{x\in\mathcal{X}}\{I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle + \langle \beta^2, \tilde{H}^2(x)\rangle\} - \varphi^1(\beta^1), \end{aligned}$$



(a) For $(\beta^1,\beta^2) \in I\!\!R^\tau \times I\!\!R^{\sigma-\tau}$, any $x \in \mathcal{E}_{\beta^1,\beta^2}$ lies in some $\mathcal{E}^{u^2}_{\beta^1}$.



(b) For $\beta^1 \in I\!\!R^\tau$, there are three possibilities for $u^2 \in \operatorname{dom} s_{\beta^1}$. The two branches on the left lead to equivalence results, whereas the other branch leads to a nonequivalence result. The sets C_{β^1} and T_{β^1} are defined in the last paragraph of Section 5.2.

Figure 2: Equivalence and nonequivalence of canonical and mixed ensembles.

where

$$\begin{aligned} \varphi^1(\beta^1) &\doteq -\lim_{n\to\infty} \frac{1}{a_n} \log \int_{\Omega^n} \exp[-a_n\langle \beta^1, H^1_n\rangle]\, dP_n \\ &= \inf_{y\in\mathcal{X}}\{I(y) + \langle \beta^1, \tilde{H}^1(y)\rangle\}. \end{aligned} \tag{5.2.4}$$

The function φ_{β^1} is finite, concave, and continuous on $I\!\!R^{\sigma-\tau}$. In (5.1.4) we identified the entropy with respect to $\text{Micro}(H^2_n; \text{Can}(H^1_n; P_n)_{a_n\beta^1})^{u^2,r}$ to be

$$\begin{aligned} s_{\beta^1}(u^2) &\doteq -J_{\beta^1}(u^2) \\ &= -\inf\{I_{\beta^1}(x) : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\} \\ &= -\inf\{I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\} + \varphi^1(\beta^1); \end{aligned} \tag{5.2.5}$$

$u^2 \in \operatorname{dom} s_{\beta^1}$ if and only if $u^2 \in \operatorname{dom} J^2$.

As in Section 4, whether or not the entropy s_{β^1} is concave on $I\!\!R^{\sigma-\tau}$, its Legendre-Fenchel transform $s^*_{\beta^1}$ equals φ_{β^1}. If in addition s_{β^1} is concave on $I\!\!R^{\sigma-\tau}$, then this formula can be inverted to give $s_{\beta^1} = \varphi^*_{\beta^1}$.

For $\beta^1 \in I\!\!R^\tau$ the relationships between $\mathcal{E}_{\beta^1,\beta^2}$ and $\mathcal{E}^{u^2}_{\beta^1}$ are summarized in Figure 2. These relationships depend on two sets that are the analogues of the sets C and T defined in (4.7) and (4.8). For $\beta^1 \in I\!\!R^\tau$ we define C_{β^1} to be the set of $u^2 \in I\!\!R^{\sigma-\tau}$ for which there exists $\beta^2 \in I\!\!R^{\sigma-\tau}$ such that

$$s_{\beta^1}(w) \leq s_{\beta^1}(u^2) + \langle \beta^2, w - u^2 \rangle \text{ for all } w \in I\!\!R^{\sigma-\tau}.$$

We also define T_{β^1} to be the set of $u^2 \in I\!\!R^{\sigma-\tau}$ for which there exists $\beta^2 \in I\!\!R^{\sigma-\tau}$ such that

$$s_{\beta^1}(w) < s_{\beta^1}(u^2) + \langle \beta^2, w - u^2 \rangle \text{ for all } w \neq u^2.$$

As in Lemma 4.1, it can be shown that $C_{\beta^1} = \Gamma_{\beta^1} \cap \operatorname{dom} \partial s^{**}_{\beta^1}$, where $\Gamma_{\beta^1} \doteq \{u^2 \in I\!\!R^{\sigma-\tau} : s_{\beta^1}(u^2) = s^{**}_{\beta^1}(u^2)\}$.

5.3 Equivalence and Nonequivalence of the Mixed and Microcanonical Ensembles

In this subsection we study, at the level of equilibrium macrostates, the equivalence and nonequivalence of the mixed ensemble $\text{Can}(H_n^1; \text{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$ and the microcanonical ensemble $\text{Micro}(H_n^1, H_n^2; P_n)^{u^1,u^2,r}$. The parameters β^1, u^1, and u^2 satisfy $\beta^1 \in I\!\!R^\tau$, $u^2 \in \operatorname{dom} J^2$, and $(u^1, u^2) \in \operatorname{dom} J$, where

$$J^2(u^2) \doteq \inf\{I(x) : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\}$$

and

$$J(u^1, u^2) \doteq \inf\{I(x) : x \in \mathcal{X}, \tilde{H}^1(x) = u^1, \tilde{H}^2(x) = u^2\}.$$

For any u^1 and u^2, $J^2(u^2) \leq J(u^1, u^2)$. Hence, if $(u^1, u^2) \in \operatorname{dom} J$, then $u^2 \in \operatorname{dom} J^2$. By a similar verification as in the paragraph after Proposition 3.1, the condition that $(u^1, u^2) \in \operatorname{dom} J$ guarantees that both the mixed ensemble and the microcanonical ensemble are well defined for all sufficiently large n. The relationships between the sets of equilibrium macrostates for the two ensembles follow immediately from Theorems 4.4, 4.6, and 4.8 with minimal changes in proof. Hence we shall only summarize them in Figure 3.

By Theorem 5.1.1, for $(\beta^1, u^2) \in I\!\!R^\tau \times (\operatorname{dom} J^2)$, with respect to $\text{Can}(H_n^1; \text{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$ Y_n satisfies the LDP with rate function

$$I^{u^2}_{\beta^1}(x) \doteq I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) + \langle \beta^1, \tilde{H}^1(x) \rangle - \psi^{u^2}_{\beta^1}, \tag{5.3.1}$$

where

$$\psi^{u^2}_{\beta^1} \doteq \inf\{I(y) + \langle \beta^1, \tilde{H}^1(y) \rangle : y \in \mathcal{X}, \tilde{H}^2(y) = u^2\}.$$

In addition, for $(u^1, u^2) \in \operatorname{dom} J$, with respect to $\text{Micro}(H_n^1, H_n^2; P_n)^{u^1,u^2,r}$ Y_n satisfies the LDP with rate function

$$I^{u^1,u^2}(x) \doteq I(\{x\} \cap (\tilde{H}^1)^{-1}(\{u^1\}) \cap (\tilde{H}^2)^{-1}(\{u^2\})) - J(u^1, u^2).$$

For $\beta^1 \in I\!\!R^\tau$, $u^2 \in \operatorname{dom} J^2$, and $(u^1, u^2) \in \operatorname{dom} J$, we define the corresponding sets of equilibrium macrostates

$$\begin{aligned} \mathcal{E}^{u^2}_{\beta^1} &\doteq \{x \in \mathcal{X} : I^{u^2}_{\beta^1}(x) = 0\} \\ &= \{x \in \mathcal{X} : \tilde{H}^2(x) = u^2, I(x) + \langle \beta^1, \tilde{H}^1(x) \rangle = \psi^{u^2}_{\beta^1}\} \end{aligned}$$

and

$$\begin{aligned}\mathcal{E}^{u^1,u^2} &\doteq \{x \in \mathcal{X} : I^{u^1,u^2}(x) = 0\} \\ &= \{x \in \mathcal{X} : I(x) = J(u^1,u^2), \tilde{H}^1(x) = u^1, \tilde{H}^2(x) = u^2\}.\end{aligned}$$

As the sets of points at which the corresponding rate functions attain their minimum of 0, the set $\mathcal{E}_{\beta^1}^{u^2}$, for $\beta^1 \in I\!R^\tau$ and $u^2 \in \operatorname{dom} J^2$, and the set $\mathcal{E}^{u^1,u^2}$, for $(u^1,u^2) \in \operatorname{dom} J$, are nonempty and compact. The purpose of this subsection is to record the relationships between these sets.

Since for $(u^1,u^2) \in \operatorname{dom} J$ and each n

$$\operatorname{Micro}(H_n^1, H_n^2; P_n)^{u^1,u^2,r} \quad \text{and} \quad \operatorname{Micro}(H_n^1; \operatorname{Micro}(H_n^2; P_n)^{u^2,r})^{u^1,r}$$

are equal, we can derive the relationships between $\mathcal{E}_{\beta^1}^{u^2}$ and $\mathcal{E}^{u^1,u^2}$ by applying the results of Section 4 to the canonical ensemble and microcanonical ensemble

$$\operatorname{Can}(H_n^1; Q_n)_{a_n\beta^2} \quad \text{and} \quad \operatorname{Micro}(H_n^1; Q_n)^{u^1,r}, \text{ with } Q_n \doteq \operatorname{Micro}(H_n^2; P_n)^{u^2,r}.$$

To this end, we introduce the relevant thermodynamic functions. By Theorem 3.2, for $u^2 \in \operatorname{dom} J^2$ the rate function in the LDP for Y_n with respect to $\operatorname{Micro}(H_n^2; P_n)^{u^2,r}$ is

$$I^{u^2}(x) \doteq I(\{x\} \cap (\tilde{H}^2)^{-1}(\{u^2\})) - J^2(u^2).$$

Hence by the Laplace principle, for $u^2 \in \operatorname{dom} J^2$ the free energy with respect to the ensemble $\operatorname{Can}(H_n^1; \operatorname{Micro}(H_n^2; P_n)^{u^2,r})_{a_n\beta^1}$ is given by

$$\begin{aligned}\varphi^{u^2}(\beta^1) &= -\lim_{n\to\infty} \frac{1}{a_n} \log \int_{\Omega^n} \exp[-a_n\langle \beta^1, H_n^1\rangle]\, d\Big(\operatorname{Micro}(H_n^2; P_n)^{u^2,r}\Big) \\ &= \inf_{x\in\mathcal{X}} \{I^{u^2}(x) + \langle \beta^1, \tilde{H}^1(x)\rangle\} \qquad (5.3.2) \\ &= \inf\{I(x) + \langle \beta^1, \tilde{H}^1(x)\rangle : x \in \mathcal{X}, \tilde{H}^2(x) = u^2\} - J^2(u^2).\end{aligned}$$

The function φ^{u^2} is finite, concave, and continuous on $I\!R^\tau$. For $u^2 \in \operatorname{dom} J^2$ we define

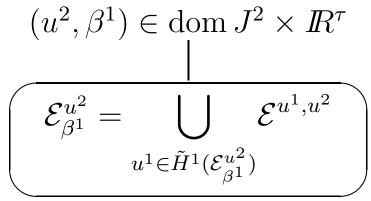
$$\begin{aligned}J^{u^2}(u^1) &\doteq \inf\{I^{u^2}(x) : x \in \mathcal{X}, \tilde{H}^1(x) = u^1\} \\ &= \inf\{I(x) : x \in \mathcal{X}, \tilde{H}^1(x) = u^1, \tilde{H}^2(x) = u^2\} - J^2(u^2) \qquad (5.3.3) \\ &= J(u^1,u^2) - J^2(u^2).\end{aligned}$$

With respect to $\operatorname{Micro}(H_n^1; \operatorname{Micro}(H_n^2; P_n)^{u^2,r})^{u^1,r}$, for $u^2 \in \operatorname{dom} J^2$ the entropy is given by

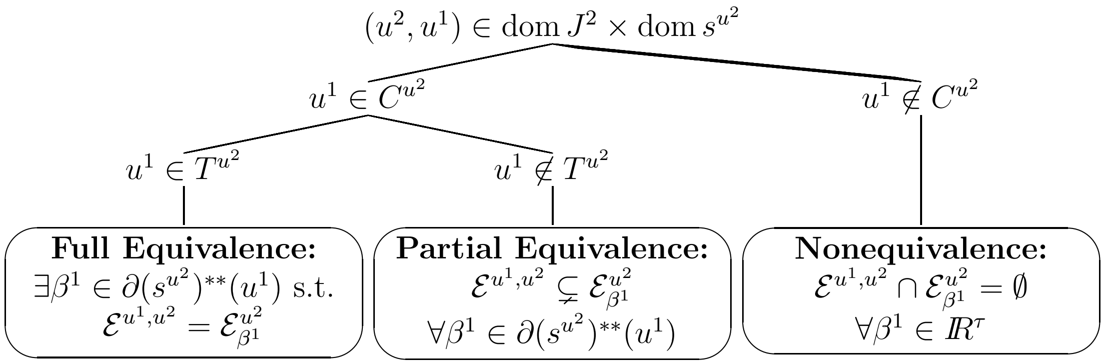
$$s^{u^2}(u^1) \doteq -J^{u^2}(u^1). \qquad (5.3.4)$$

We have $u^1 \in \operatorname{dom} s^{u^2}$ if and only if $(u^1,u^2) \in \operatorname{dom} J$.

As in Section 4, whether or not s^{u^2} is concave on $I\!R^\tau$, its Legendre-Fenchel transform $(s^{u^2})^*$ equals φ^{u^2}. If s^{u^2} is concave on $I\!R^\tau$, then this formula can be inverted to give $s^{u^2} = (\varphi^{u^2})^*$ for all $u^1 \in I\!R^\tau$.



(a) For $(u^2, \beta^1) \in \operatorname{dom} J^2 \times I\!\!R^\tau$, any $x \in \mathcal{E}^{u^2}_{\beta^1}$ lies in some $\mathcal{E}^{u^1,u^2}$.



(b) For $u^2 \in \operatorname{dom} J^2$, there are three possibilities for $u^1 \in \operatorname{dom} s^{u^2}$. The two branches on the left lead to equivalence results, whereas the other branch leads to a nonequivalence result. The sets C^{u^2} and T^{u^2} are defined in the next to last paragraph of Section 5.3.

Figure 3: Equivalence and nonequivalence of mixed and microcanonical ensembles.

For $u^2 \in \operatorname{dom} J^2$ the relationships between $\mathcal{E}^{u^2}_{\beta^1}$ and $\mathcal{E}^{u^1,u^2}$ are summarized in Figure 3. These relationships depend on two sets that are the analogues of the sets C and T defined in (4.7) and (4.8). For $\beta^1 \in I\!\!R^\tau$ we define C^{u^2} to be the set of $u^1 \in I\!\!R^\tau$ for which there exists $\beta^1 \in I\!\!R^\tau$ such that

$$s^{u^2}(w) \leq s^{u^2}(u^1) + \langle \beta^1, w - u^1 \rangle \text{ for all } w \in I\!\!R^\tau.$$

We also define T^{u^2} to be the set of $u^1 \in I\!\!R^\tau$ for which there exists $\beta^1 \in I\!\!R^\tau$ such that

$$s^{u^2}(w) < s^{u^2}(u^1) + \langle \beta^1, w - u^1 \rangle \text{ for all } w \neq u^1.$$

As in Lemma 4.1, it can be shown that $C^{u^2} = \Gamma^{u^2} \cap \operatorname{dom} \partial(s^{u^2})^{**}$, where $\Gamma^{u^2} \doteq \{u^1 \in I\!\!R^\tau : s^{u^2}(u^1) = (s^{u^2})^{**}(u^1)\}$.

With Figure 3, we complete our presentation of the equivalence and nonequivalence results for the mixed ensemble, the canonical ensemble, and the microcanonical ensemble.

References

[1] M. Aizenman, S. Goldstein, J. L. Lebowitz. "Conditional equilibrium and the equivalence of microcanonical and grand canonical ensembles in the thermodynamic limit." *Commun. Math. Phys.* 62:279-302, 1978.

[2] V. I. Arnold. *Mathematical Methods of Classical Mechanics.* Trans. K. Vogtmann and A. Weinstein. New York: Springer-Verlag, 1978.

[3] V. I. Arnold and B. A. Khesin. *Topological Methods in Hydrodynamics.* New York: Springer-Verlag, 1998.

[4] R. Balescu. *Equilibrium and Nonequilibrium Statistical Mechanics* New York: John Wiley & Sons, 1975.

[5] T. Bodineau and A. Guionnet. "About the stationary states of vortex systems." *Ann. Inst. Henri Poincaré* 35:205-237, 1999.

[6] C. Boucher, R. S. Ellis, and B. Turkington. "Derivation of maximum entropy principles in two-dimensional turbulence via large deviations." *J. Statist. Phys.* 98:1235-1278, 2000.

[7] C. Boucher, R. S. Ellis, and B. Turkington. "Spatializing random measures: doubly indexed processes and the large deviation principle." *Ann. Probab.* 27:297-324, 1999.

[8] E. Caglioti, P. L. Lions, C. Marchioro, and M. Pulvirenti. "A special class of stationary flows for two-dimensional Euler equations: a statistical mechanical description." *Commun. Math. Phys.* 143:501-525 (1992).

[9] E. Caglioti, P. L. Lions, C. Marchioro, and M. Pulvirenti. "A special class of stationary flows for two-dimensional Euler equations: a statistical mechanical description. Part II." *Commun. Math. Phys.* 174:229-260 (1995).

[10] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Second edition. New York: Springer-Verlag, 1998.

[11] J.-D. Deuschel and D. W. Stroock. *Large Deviations.* Boston: Academic Press, 1989.

[12] J.-D. Deuschel, D. W. Stroock, and H. Zessin. "Microcanonical distributions for lattice gases." *Commun. Math. Phys.* 139:83-101, 1991.

[13] M. DiBattista and A. Majda. "An equilibrium statistical theory for large-scale features of open-ocean convection." To appear in *J. Phys. Oceanography*, 2000.

[14] M. DiBattista, A. Majda, and B. Turkington. "Prototype geophysical vortex structures via large-scale statistical theory." *Geophys. Astrophys. Fluid Dyn.* 89:235-283, 1998.

[15] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* New York: John Wiley & Sons, 1997.

[16] R. S. Ellis, K. Haven, and B. Turkington. "Nonequivalent statistical equilibrium ensembles and refined stability theorems for most probable flows." 41 page Latex manuscript. Submitted for publication, 2000.

[17] R. S. Ellis, R. Jordan, and B. Turkington. "A large deviation approach to soliton turbulence for the nonlinear Schrödinger equation." In preparation, 2000.

[18] R. E. Ellis and J. S. Rosen. "Laplace's method for Gaussian integrals with an application to statistical mechanics." *Ann. Probab.* 10:47-66, 1982. Correction in *Ann. Probab.* 11:456, 1983.

[19] G. L. Eyink and H. Spohn. "Negative-temperature states and large-scale, long-lived vortices in two-dimensional turbulence." *J. Statist. Phys.* 70:833-886, 1993.

[20] H. Föllmer and S. Orey. "Large deviations for the empirical field of a Gibbs measure." *Ann. Probab.* 16:961-977, 1987.

[21] H.-O. Georgii. "Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$. *Ann. Probab.* 21:1845-1875, 1993.

[22] A. D. Ioffe and V. M. Tihomirov. *Theory of Extremal Problems.* New York: Elsevier North-Holland, 1979.

[23] R. Jordan and C. Josserand. "Self-organization in nonlinear wave turbulence." *Phys. Rev. E* 61:1527-1539, 2000.

[24] R. Jordan, B. Turkington, and C. L. Zirbel. "A mean-field statistical theory for the nonlinear Schrödinger equation." *Physica D* 137:353-378, 2000.

[25] G. Joyce and D. C. Montgomery. "Negative temperature states for the two-dimensional guiding center plasma." *J. Plasma Phys.* 10:107-121, 1973.

[26] M. K.-H. Kiessling. "Statistical mechanics of classical particles with logarithmic interactions." *Commun. Pure Appl. Math.* 46:27-56, 1993.

[27] M. K.-H. Kiessling. "Statistical Mechanics of Weakly Dissipative Current-Carrying Plasma." Habilitationsschrift, Ruhr-Universität Bochum, 1995.

[28] M. K.-H. Kiessling and Joel L. Lebowitz. "The micro-canonical point vortex ensemble: beyond equivalence." *Lett. Math. Phys.* 42:43-56, 1997.

[29] M. K.-H. Kiessling and T. Neukirch. "Negative specific heat of toroidally confined plasma." Technical report, 1997.

[30] R. Kraichnan. "Statistical dynamics of two-dimensional flow." *J. Fluid Mech.* 67:155-175, 1975.

[31] O. E. Lanford. "Entropy and equilibrium states in classical statistical mechanics." *Statistical Mechanics and Mathematical Problems.* Ed. A. Lenard. Lecture Notes in Physics, Volume 20 Berlin: Springer–Verlag, 1973. 1-113.

[32] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan. "Entropy, concentration of probability and conditional limit theorems." *Markov Proc. Related Fields* 1:319-386, 1995.

[33] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan. "The equivalence of ensembles for lattice systems: some examples and a counterexample." *J. Statist. Phys.* 77:397-419 (1994).

[34] J. Lynch and J. Sethuraman. "Large deviations for processes with independent increments." *Ann. Probab.* 15:610-627, 1987.

[35] C. Marchioro and M. Pulvirenti. *Mathematical Theory of Incompressible Nonviscous Fluids*. New York: Springer-Verlag, 1994.

[36] J. Messer and H. Spohn. "Statistical mechanics of the Lane-Emden equation." *J. Statist. Phys.* 29:561-578, 1982.

[37] J. Michel and R. Robert. "Large deviations for Young measures and statistical mechanics of infinite dimensional dynamical systems with conservation law." *Commun. Math. Phys.* 159:195-215, 1994.

[38] J. Miller. "Statistical mechanics of Euler equations in two dimensions." *Phys. Rev. Lett.* 65:2137-2140, 1990.

[39] J. Miller, P. Weichman, and M. C. Cross. "Statistical mechanics, Euler's equations, and Jupiter's red spot." *Phys. Rev. A* 45:2328-2359, 1992.

[40] S. Olla. "Large deviations for Gibbs random fields." *Probab. Th. Related Fields* 77:343-359, 1988.

[41] L. Onsager. "Statistical hydrodynamics." *Suppl. Nuovo Cim.* 6:279-287, 1949.

[42] R. Robert. "Concentration et entropie pour les mesures d'Young." *C. R. Acad. Sci. Paris*, Série I, 309:757-760, 1989.

[43] R. Robert. "A maximum-entropy principle for two-dimensional perfect fluid dynamics." *J. Statist. Phys.* 65:531-553, 1991.

[44] R. Robert and J. Sommeria. "Statistical equilibrium states for two-dimensional flows." *J. Fluid Mech.* 229:291-310, 1991.

[45] R. T. Rockafellar. *Convex Analysis*. Princeton: Princeton University Press, 1970.

[46] S. Roelly and H. Zessin. "The equivalence of equilibrium principles in statistical mechanics and some applications to large particle systems." *Expositiones Mathematicae* 11:384-405, 1993.

[47] J. Schröter and R. Wegener. "The problem of equivalence in statistical mechanics of equilibrium." *Math. Methods Appl. Sci.* 14:319-331, 1991.

[48] R. A. Smith and T. M. O'Neil. "Nonaxisymmetric thermal equilibria of a cylindrically bounded guiding center plasma or discrete vortex system." *Phys. Fluids B* 2:2961-2975, 1990.

[49] W. Thirring. *A Course in Mathematical Physics 4: Quantum Mechanics of Large Systems*. Trans. E. M. Harrell. New York: Springer-Verlag, 1983.

[50] B. Turkington. "Statistical equilibrium measures and coherent states in two-dimensional turbulence." *Commun. Pure Appl. Math* 52:781-809, 1999.

[51] B. Turkington and N. Whitaker. "Statistical equilibrium computations of coherent structures in turbulent shear layers." *SIAM J. Sci. Comput.* 17:1414-1433, 1996.

[52] N. Whitaker and B. Turkington. "Maximum entropy states for rotating vortex patches." *Phys. Fluids A* 6:3963-3973, 1994.

[53] E. Zeidler. *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*. New York: Springer-Verlag, 1985.