Journal of Statistical Physics, Vol. 98, Nos. 5/6, 2000

Derivation of Maximum Entropy Principles in Two-Dimensional Turbulence via Large Deviations

Christopher Boucher,[1] Richard S. Ellis,[2] and Bruce Turkington[3]

Received March 4, 1999; final October 14, 1999

The continuum limit of lattice models arising in two-dimensional turbulence is analyzed by means of the theory of large deviations. In particular, the Miller–Robert continuum model of equilibrium states in an ideal fluid and a modification of that model due to Turkington are examined in a unified framework, and the maximum entropy principles that govern these models are rigorously derived by a new method. In this method, a doubly indexed, measure-valued random process is introduced to represent the coarse-grained vorticity field. The natural large deviation principle for this process is established and is then used to derive the equilibrium conditions satisfied by the most probable macrostates in the continuum models. The physical implications of these results are discussed, and some modeling issues of importance to the theory of long-lived, large-scale coherent vortices in turbulent flows are clarified.

KEY WORDS: fluid turbulence, statistical equilibrium, large deviation principles.

1. INTRODUCTION

Both physical observations and numerical simulations show that a freely evolving, turbulent fluid in two dimensions tends to form long-lived, large-scale coherent structures.[26, 36] If the flow is allowed to evolve for a sufficiently long time, these structures organize into a steady, stable vortex or shear flow that persists within the small-scale turbulent fluctuations of the vorticity field.[31, 37] While this scenario pertains to the idealized limit of

[1] Department of Mathematics, Illinois Wesleyan University, Bloomington, Illinois 61702; e-mail: cboucher@sun.iwu.edu.

[2] Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003; e-mail: rsellis@math.umass.edu.

[3] Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003; e-mail: turk@math.umass.edu.

0022-4715/00/0300-1235$18.00/0

Navier–Stokes equations in which the Reynolds number is sent to infinity, it is nevertheless a distinguishing feature of turbulent two-dimensional fluid motions that contrasts sharply with the general behavior of fully three-dimensional flows.[6]

In this paper we study a class of models that have been proposed to elucidate the self-organization of these coherent states. The continuum models with which we are concerned are governed by maximum entropy principles and are designed to distinguish the most probable macroscopic states among all possible such states. These principles are, in turn, derived from a microscopic model of the random vorticity field via the methodology of equilibrium statistical mechanics. Indeed, the equations of motion of an ideal fluid constitute a Hamiltonian system with infinitely many degrees of freedom, for which, in principle, the standard methods of statistical equilibrium theory are applicable. The construction of a meaningful Gibbs measure, however, requires that the continuum system be replaced by an approximating system with finitely many degrees of freedom and that an appropriately scaled continuum limit be taken. It is precisely on this issue that the present paper is focused. Specifically, we examine in detail how to employ the theory of large deviations to derive the maximum entropy principles that govern these continuum models.

Such a continuum model of coherent structures has been proposed independently by Miller *et al.*[28, 29] and Robert *et al.*,[34, 35] and we shall refer to it as the Miller–Robert model. This theory builds on earlier and simpler theories, which generally are derived by two different approaches. In one approach, initiated by Onsager[32] and later developed by Joyce and Montgomery[21, 30] and others,[17, 22, 4] the fluid vorticity is replaced by a dilute gas of point vortices, and a maximum entropy principle is obtained from a mean-field theory of the point vortex system. In the other approach, investigated by Kraichnan[23] and others,[3] the vorticity is represented by a finite number of Fourier modes, a spectrally projected dynamics is defined, and a canonical ensemble is constructed from the two quadratic invariants for these dynamics. Both of these approaches suffer from the defect that they retain only a few of the dynamical constraints inherent in the underlying partial differential equations. For this reason, information contained in the continuum dynamics is lost in passing to a finite-dimensional, approximate dynamics and then to a continuum limit. The important innovation of the Miller–Robert model is that it incorporates all the conserved quantities of two-dimensional ideal fluid dynamics. Specifically, the Miller–Robert model includes the entire family of global vorticity integrals, also known as generalized enstrophies,[18] along with the energy integral as constraints in its maximum entropy principle. It therefore subsumes the earlier theories as special cases or as limiting cases.

Nevertheless, the Miller–Robert model is based on some implicit assumptions about the structure of small-scale fluctuations of vorticity, which are not supported by the underlying continuum dynamics. Unlike its simpler predecessors, the model is not derived from a sequence of finite-dimensional approximations, each having a well-defined dynamics and corresponding conserved quantities. Instead, the theory is deduced by postulating that the discrete analogue of the conserved quantities for the continuum dynamics can simply be imposed on each finite-dimensional statistical equilibrium model. Recently, Turkington[(39)] has criticized these assumptions and has proposed a modification of the Miller–Robert model that is based on invariance with respect to the exact continuum dynamics. In this paper we shall refer to this modification as Turkington's model. The essential difference between these two models, which otherwise have a similar structure, lies in the way that the global vorticity integrals enter as constraints into their corresponding maximum entropy principles. As a consequence of this difference in the microscopic modeling, the two theories make different macroscopic predictions about coherent states.

Rather than enter into modeling issues, in the present paper we shall concentrate on showing how models of this kind can be analyzed with the same machinery as is employed to study other models in statistical mechanics, such as standard models of ferromagnetism. In particular, we shall demonstate how these models of coherent states in two-dimensional turbulence fit into the general framework articulated by Ellis for the large deviation analysis of statistical equilibrium models (ref. 15, Sections 9–12). In this context, we shall obtain our main result, which is a new, rigorous derivation of the continuum limit for the Miller–Robert model and Turkington's model. A key step in this analysis is the identification of a so-called "hidden process" that links the microscopic model, defined by a sequence of lattice Gibbs measures, to the macroscopic continuum model, governed by a variational principle involving a relative entropy and appropriate representations of the invariants that enter into the Gibbs measures. With respect to the Gibbs measures, this hidden process satisfies a large deviation principle, which distinguishes the equilibrium macrostates as the most probable states in the sense that any other macrostate has an exponentially small probability of being observed. The variational principle for the equilibrium macrostates is constructed from the rate function for this large deviation principle.

With this general approach we extend and clarify, both from a conceptual viewpoint and a technical viewpoint, the program initiated by Miller, Robert, and co-workers. From a conceptual viewpoint, we remove an element of doubt that has existed about the validity of these models. The doubt concerns the question of whether the continuum models and their maximum entropy principles are entirely determined by taking the limit of some well-defined Gibbs measures or whether they are mean-field theories that rely on some additional approximation. This question has been raised both in theoretical attempts to justify the Miller–Robert model[17] and in numerical simulations to validate its predictions.[6] We show that the continuum limit of such models, which is scaled differently from the usual thermodynamic limit, is indeed exactly described by their maximum entropy principles. We are therefore able to say that the attributes of models of this kind stem from how their finite-dimensional Gibbs measures capture the underlying continuum dynamics. We illustrate this point by contrasting the formulation of Miller–Robert with the modification due to Turkington.

From a technical viewpoint, we offer a unified and rigorous method of analyzing the continuum limits for both the Miller–Robert model and Turkington's model. This method is the mathematical expression of the physical reasoning pioneered by Miller *et al.*,[28, 29] which focuses on the canonical ensemble. Our key innovations are, first, to introduce a certain doubly indexed process that approximates the hidden process and, second, to derive the large deviation principle for the hidden process from a large deviation principle for the doubly indexed process. We construct this doubly indexed approximation by averaging over an intermediate scale, thereby capturing in a natural way the coarse-graining inherent in the macrostates. Not only does our construction embody the physical insights in the work of Miller *et al.*,[28, 29] but also it allows us to derive the continuum limits directly from the familiar Sanov's Theorem for the empirical measures of i.i.d. random variables. On the other hand, Michel and Robert base their large deviation analysis of the continuum limit of the Miller–Robert model on the microcanonical ensemble,[27, 33] using an indirect argument that relies on Baldi's Theorem, an infinite-dimensional generalization of the Gärtner-Ellis Theorem (ref. 9, Thm 4.5.20). In contrast to our approach, theirs is less closely tied to the underlying physics and, in particular, de-emphasizes the role of spatial coarse-graining in the large deviation behavior.

A number of issues related to the present paper are treated in some detail in a sequel.[16] There, we study the continuum limit of the microcanonical ensemble for a class of statistical mechanical models that includes the Miller–Robert model and Turkington's model as special cases. In the context of the equivalence of the microcanonical and canonical ensembles, we also compare those results with the results obtained in the present paper.

The theorems proved here may have applications beyond the realm of ideal equilibrium theories. For instance, the recent work of Chorin, Kast, and Kupferman[8] suggests that equilibrium measures may be used to approximate the evolution of a system with many degrees of freedom by building mean equations for a few well chosen observables. These ideas open some exciting possibilities for modeling the temporal behavior of two-dimensional turbulence.

The paper is organized as follows. In Section 2, we formulate the lattice models of Miller–Robert and of Turkington in a unified fashion, and we indicate the modeling hypotheses used in their constructions. We state our main theorem in Subsection 3.1 [Theorem 3.1] after laying out the general framework for the large deviation analysis. Then in Subsections 3.2 and 3.3 we pause to discuss the fluid dynamical implications of our main results and the maximum entropy principles for the two continuum models. In Section 4, we complete the proof of the main theorem by establishing the properties of the hidden process that are needed to invoke our general method; here we define the doubly indexed process and prove that it approximates the hidden process. Finally, in Section 5 we outline the proof of the large deviation principle for the doubly indexed process itself. This result is a special case of a general theorem that we give in another paper.[1]

2. STATISTICAL EQUILIBRIUM MODELS

2.1. Gibbs Measures and Lattice Models

In two dimensions, the equations governing the evolution of an ideal fluid are reducible to the vorticity transport equation[6, 25]

$$\frac{\partial \omega}{\partial t}+\frac{\partial \omega}{\partial x_1}\frac{\partial \psi}{\partial x_2}-\frac{\partial \omega}{\partial x_2}\frac{\partial \psi}{\partial x_1}=0, \qquad -\Delta\psi=\omega \tag{2.1}$$

in which ω is the vorticity, ψ is the stream function, and $\Delta = \partial^2/\partial x_1^2 + \partial/\partial x_2^2$ denotes the Laplacian operator on $\mathbb{R}^2$. The two-dimensionality of the flow means that these quantities are related to the velocity field $v = (v_1, v_2, 0)$ according to $(0, 0, \omega) = \text{curl}\, v$ and $v = \text{curl}(0, 0, \psi)$. All of these fields depend upon the time variable $t \in [0, \infty)$ and the space variable $x = (x_1, x_2)$, which runs through a bounded domain $\mathcal{X} \subset \mathbb{R}^2$. Throughout this paper we assume that $\mathcal{X}$ equals the unit torus $T^2 \doteq [0, 1] \times [0, 1]$ equipped with doubly periodic boundary conditions. By making this technical simplification, we follow a common practice in most theoretical and numerical studies of two-dimensional turbulence. Nevertheless, our results have natural extensions from this prototype geometry to general bounded domains with physically appropriate boundary conditions.

The governing equations (2.1), which are equivalent to the Euler equations for an incompressible, inviscid fluid, can be expressed as a single equation for the scalar vorticity field $\omega = \omega(x, t)$. In our prototype geometry where $\mathscr{X} = T^2$, the periodicity of the velocity field implies that $\int_{\mathscr{X}} \omega \, dx = 0$. With this restriction on its domain, the Green's operator $G = (-\Delta)^{-1}$ taking ω into ψ with $\int_{\mathscr{X}} \psi \, dx = 0$ is well-defined. More explicitly, G is the integral operator

$$\psi(x) = G\omega(x) = \int_{\mathscr{X}} g(x - x')\, \omega(x')\, dx' \tag{2.2}$$

where g is the Green's function defined by the Fourier series

$$g(x - x') \doteq \sum_{0 \neq z \in \mathbb{Z}^2} |2\pi z|^{-2}\, e^{2\pi i z \cdot (x - x')} \tag{2.3}$$

Consequently, (2.1) can be considered as an equation in ω alone. The vorticity transport equation (2.1) has a natural interpretation in terms of the flow maps $\phi^t \colon \mathscr{X} \to \mathscr{X}$ defined by fluid particle trajectories $x(t) = \phi^t(x^0)$, where $dx/dt = v(x, t)$ and $x(0) = x^0$. Namely, $\omega(\phi^t(x^0), t) = \omega^0(x^0)$ for all $x^0 \in \mathscr{X}$ and for all $t > 0$. Thus, the vorticity is rearranged by the area-preserving flow maps for the velocity field, which itself is induced by the vorticity at each instant of time.[6, 25]

Even though the initial value problem for the equation (2.1) is known to be well-posed for weak solutions whenever the initial data $\omega^0 = \omega(\cdot, 0)$ belongs to $L^\infty(\mathscr{X})$,[25] this determinstic evolution does not provide a useful description of the system over long time intervals. Indeed, as many numerical simulations and physical experiments show, a typical solution develops finite amplitude fluctuations in its vorticity field on increasingly fine spatial scales as time proceeds. These fluctuations result from the self-straining of the vorticity by the induced flow maps ϕ^t, which engenders a rapid growth in the vorticity gradient $|\nabla\omega|$. For this reason, the information required to specify the deterministic state ω exceeds any given bound after a sufficiently long time. When one seeks to quantify the long-time behavior of solutions, therefore, one is compelled to shift from the microscopic, or fine-grained, description inherent in ω to some kind of macroscopic, or coarse-grained, description. We shall make this shift by adopting the perspective of equilibrium statistical mechanics. That is, we shall view the underlying deterministic dynamics as a means of randomizing the microstate ω subject to the conditioning inherent in the conserved quantities for the governing equation (2.1), and we shall take the appropriate macrostates to be the canonical Gibbs measures $P(d\omega)$ built from these conserved quantities. In doing so, we accept an ergodic hypothesis that equates the time averages with canonical ensemble averages. Given this hypothesis, we expect that these macrostates capture the long-lived, large-scale, coherent vortex structures that persist among the small-scale vorticity fluctuations. The characterization of these self-organized macrostates, which are observed in simulations and experiments, is the ultimate goal of our theory.

The conserved quantities that determine the Gibbs state are the energy, or Hamiltonian functional, and the family of generalized enstrophies, or Casimir functionals.[25] Expressed as a functional of ω, the kinetic energy is

$$H(\omega) \doteq \frac{1}{2} \int_{\mathscr{X} \times \mathscr{X}} g(x-x')\, \omega(x)\, \omega(x')\, dx\, dx' \tag{2.4}$$

The so-called generalized enstrophies are the global vorticity integrals

$$A(\omega) \doteq \int_{\mathscr{X}} a(\omega(x))\, dx \tag{2.5}$$

where a is an arbitrary continuous real-valued function on the range of the vorticity. These additional invariants A arise from the fact that the vorticity is rearranged under the area-preserving flow maps ϕ^t. Their presence in two-dimensional vortex dynamics gives the statistical equilibrium theory its distinctive features.

In terms of the dynamical invariants H and A, the canonical ensemble is defined by the formal Gibbs measure

$$P_{\beta, a}(d\omega) = Z(\beta, a)^{-1} \exp[-\beta H(\omega) - A(\omega)]\; \Pi(d\omega) \tag{2.6}$$

where $Z(\beta, a)$ is the associated partition function and $\Pi(d\omega)$ denotes some invariant product measure on some phase space of all admissible vorticity fields ω. Of course, this formal construction is not meaningful as it stands, due to the infinite dimensionality of such a phase space. We therefore proceed to define a sequence of lattice models on $\mathscr{X} = T^2$ in order to give a meaning to this formal construction.

Let $\mathscr{L}$ be a uniform lattice of $n \doteq 2^{2m}$ sites s in the unit torus T^2, where m is a positive integer. The intersite spacing in each coordinate direction is 2^{-m}. We make this particular choice of n to ensure that the lattices are refined dyadically as m increases, a property that is needed later when we study the continuum limit obtained by sending $n \to \infty$ along the sequence $n = 2^{2m}$. Each such lattice of n sites induces a dyadic partition of T^2 into n squares called microcells, each having area $1/n$. For each $s \in \mathscr{L}$ we denote by $M(s)$ the unique microcell containing the site s in its lower left corner. Although $\mathscr{L}$ and $M(s)$ depend on n, this is not indicated in the notation.

The phase space for the lattice models is the product space $\Omega_n \doteq \mathscr{Y}^n$, where $\mathscr{Y}$ is a compact set in $\mathbb{R}$ with $K \doteq \max\{|y|: y \in \mathscr{Y}\}$. The configurations in Ω_n are the microstates for the lattice model on $\mathscr{L}$ and are denoted by $\zeta = \{\zeta(s), s \in \mathscr{L}\}$. These microstates ζ represent a discretization of the continuum vorticity field $\omega \in L^\infty(\mathscr{X})$. Accordingly, $\mathscr{Y}$ is taken to be a compact set containing the range of the vorticity, which is an invariant of the governing equation (2.1). We let $\mathscr{P}(\Omega_n)$ denote the set of Borel probability measures on Ω_n.

On the phase space Ω_n we consider a finite product measure Π_n that assigns to a Borel subset B of Ω_n the probability

$$\Pi_n\{B\} \doteq \int_B \prod_{s \in \mathscr{L}} \rho(d\zeta(s)) \tag{2.7}$$

In this formula ρ is a given probability measure on $\mathscr{Y}$ that is meant to encode the prior information on the microscopic fluctuations in the lattice model. With respect to Π_n, $\{\zeta(s), s \in \mathscr{L}\}$ is an independent, identically distributed collection of random variables with common distribution ρ. We refer to ρ as the prior distribution. The prior distribution ρ and its support $\mathscr{Y}$ are particular ingredients in each lattice model, and their choice is a modeling issue that requires separate justification. For this reason, we defer the complete specification of ρ and $\mathscr{Y}$ until the later subsections where we describe two particular models individually.

The Hamiltonian for the lattice model is defined in terms of the spectral truncation of the Green's function g introduced in (2.3). The lattice Hamiltonian H_n maps Ω_n into $\mathbb{R}$ and is defined by

$$H_n(\zeta) \doteq \frac{1}{2n^2} \sum_{s, s' \in \mathscr{L}} g_n(s-s')\, \zeta(s)\, \zeta(s') \tag{2.8}$$

where g_n is the lattice Green's function defined by the finite Fourier sum

$$g_n(s-s') \doteq \sum_{0 \neq z \in \mathscr{L}^*} |2\pi z|^{-2}\, e^{2\pi i z \cdot (s-s')} \tag{2.9}$$

over the finite set

$$\mathscr{L}^* \doteq \{z = (z_1, z_2) \in \mathbb{Z}^2 : -2^{m-1} < z_1, z_2 \leqslant 2^{m-1}\}$$

The generalized enstrophy A_n for the lattice model maps Ω_n into $\mathbb{R}$ and is defined by

$$A_n(\zeta) \doteq \frac{1}{n} \sum_{s \in \mathscr{L}} a(\zeta(s)) \tag{2.10}$$

where a is any continuous function mapping $\mathscr{Y}$ into $\mathbb{R}$. Although A_n depends on a, this is not indicated in the notation. The functions H_n and A_n on Ω_n are the natural discretizations of the corresponding functionals of the vorticity field given in (2.4) and (2.5), respectively.

In terms of these quantities we define the partition function

$$Z(n, \beta, a) \doteq \int_{\Omega_n} \exp[-\beta H_n(\zeta) - A_n(\zeta)]\, \Pi_n(d\zeta) \tag{2.11}$$

and the lattice Gibbs state $P_{n, \beta, a}$, which is the probability measure that assigns to a Borel subset B of Ω_n the probability

$$P_{n, \beta, a}\{B\} \doteq \frac{1}{Z(n, \beta, a)} \int_B \exp[-\beta H_n(\zeta) - A_n(\zeta)]\, \Pi_n(d\zeta) \tag{2.12}$$

These probability measures are parametrized by the constant $\beta \in \mathbb{R}$ and the function $a \in C(\mathscr{Y})$. The dependence of Gibbs measures on the inverse temperature β is standard, while their dependence on the function a that determines the enstrophy functional is a novelty of this particular statistical equilibrium problem.

At this point in the formulation of the statistical equilibrium theory, it is necessary to connect the parameters β and a in the lattice Gibbs states with the global conserved quantities to which they correspond. Indeed, the probability measure $P_{n, \beta, a}$ in (2.12) is intended to describe the time-averaged behavior of the ideal fluid flow, which is assumed to evolve ergodically from a given initial state ω^0. The problem of interest, therefore, is to determine β and a so that the ensemble mean values $\langle H_n \rangle$ and $\langle A_n \rangle$ coincide with the values derived from an initial state. While the duality between β and $\langle H_n \rangle$ is standard, the determination of a from the family of generalized enstrophy mean values is a subtle point that requires careful attention. Moreover, unlike the Hamiltonian H_n, the generalized enstrophy A_n is sensitive to the smallest scales on which vorticity fluctuations occur. Accordingly, the inclusion of A_n in the lattice Gibbs state is coupled with the choice of the prior distribution ρ, which controls the microstate fluctuations in the lattice model. In the next two subsections, therefore, we discuss this point explicitly for the two different models—the one due to Miller[28] and Robert[34] and the one proposed by Turkington.[39] These models are derived from different hypotheses about the microscopic properties of typical ideal fluid motions.

2.2. Miller–Robert Model

The theory proposed by Miller[28] and Robert[34] is based on a direct transcription of the invariants from the continuum dynamics to the lattice model. In other words, it does not construct a lattice dynamics and then identify the conserved quantities for this dynamics, but instead imposes a discretized version of the continuum invariants on the lattice model. We can summarize their approach as follows.

We seek to include the complete family of generalized enstrophy invariants for ideal fluid flow in $\mathcal{X}$ into the model; namely,

$$\int_{\mathcal{X}} a(\omega(x,t))\,dx = \int_{\mathcal{X}} a(\omega^0(x))\,dx \qquad \text{for all} \quad t>0$$

where $\omega(x,t)$ is the solution to (2.1) with initial condition $\omega^0(x)$. Since a is an arbitrary continuous function, these invariants can be rewritten in the measure-valued form

$$\int_{\mathcal{X}} \delta_{\omega(x,t)}(dy)\,dx = \int_{\mathcal{X}} \delta_{\omega^0(x)}(dy)\,dx \qquad \text{for all} \quad t>0$$

On this basis we choose the measure ρ that defines the finite product measure (2.7) to be

$$\rho(dy) \doteq \int_{\mathcal{X}} \delta_{\omega^0(x)}(dy)\,dx \tag{2.13}$$

Thus, the prior distribution ρ for the lattice model is taken to be the exact invariant for the continuum dynamics, which is fixed by the initial state ω^0. Also, we let $\mathcal{Y}$ be the support of ρ, which coincides with the closure of the range of ω^0. In this way, the exact rearrangement of vorticity ω that holds for the continuum dynamics is imposed on the microstates ζ for the lattice model.

The function a that appears in the lattice Gibbs state $P_{n,\beta,a}$ is taken to be dual to the constraint

$$\left\langle \frac{1}{n} \sum_{s \in \mathcal{L}} \delta_{\zeta(s)}(dy) \right\rangle = \rho(dy) \tag{2.14}$$

in the same way that β is dual to the constraint

$$\langle H_n(\zeta) \rangle = H(\omega^0) \tag{2.15}$$

Here, $\langle \cdot \rangle$ denotes the ensemble average defined by $P_{n,\beta,a}$. In other words, given an initial state ω^0 and any $n = 2^{2m}$, we adjust the parameters β and a so that the constraints (2.14) and (2.15) are fulfilled. Formally, one can see that β and a arise as Lagrange multipliers associated with these constraints by casting the lattice Gibbs measure as the solution to the constrained minimization problem

$$\text{minimize } R(P \mid \Pi_n) \text{ over } P \in \mathscr{P}(\Omega_n) \text{ subject to}$$

$$\int_{\Omega_n} H_n(\zeta)\, P(d\zeta) = H(\omega^0), \qquad \int_{\Omega_n} \left(\frac{1}{n} \sum_{s \in \mathscr{L}} \delta_{\zeta(s)}(dy) \right) P(d\zeta) = \int_{\mathscr{X}} \delta_{\omega^0(x)}(dy)\, dx \tag{2.16}$$

in which the objective functional is the relative entropy

$$R(P \mid \Pi_n) \doteq \begin{cases} \int_{\Omega_n} \left(\log \dfrac{dP}{d\Pi_n}(\zeta) \right) P(d\zeta) & \text{if} \quad P \ll \Pi_n \\ \infty & \text{otherwise} \end{cases} \tag{2.17}$$

The solution $P = P_{n, \beta, a}$ to this problem exists and is unique because the objective functional is strictly convex and lower semicontinous with respect to the weak convergence of measures and the constraints are imposed on linear functionals. A formal calculation with the Lagrange multiplier rule shows that this solution is the lattice Gibbs state (2.12), in which β and a, respectively, are multipliers for the constraints on energy and global vorticity distribution.[19, 42]

In this way the statistical equilibrium theory is closed, and the lattice model is completely specified in the sense that the organized states, or coherent structures, which are expected to emerge from a given initial state ω^0 are identified with that lattice Gibbs measure whose defining parameters are determined by ω^0. Through the values of its energy and generalized enstrophy, the initial state supplies the prescribed, or controllable, parameters in the theory, while the ensemble parameters β and a are determined by these values.

2.3. Turkington's Model

The implicit assumption in the Miller–Robert theory that the microstates ζ on the lattice $\mathscr{L}$ satisfy the same constraints as the solutions ω to the Euler equations on the domain $\mathscr{X}$ is criticized in ref. 39, where a modified theory based on more realistic assumptions is developed. We now summarize the motivation for this alternative model, referring the reader to ref. 39 for a full discussion and justification.

We seek a statistical equilibrium theory that is consistent with the exact continuum dynamics (2.1). In formulating a lattice model, therefore, we need to determine how the conservation properties of the continuum vorticity fields ω are reflected in the lattice microstates ζ. For any finite n, these microstates ζ can be viewed as local averages of the fields ω, and under such an averaging the continuum dynamics on $\mathscr{X}$ can be observed on the lattice $\mathscr{L}$. In ref. 39, this approach is carried out using a spatial averaging, or a spectral windowing, defined by convolutions $\zeta = K_n * \omega$ with positive approximate identities K_n chosen to scale with the lattice spacing 2^{-m} as $n = 2^{2m} \to \infty$; that is, the limit of $2^{-2m} K_n(2^{-m}x)$ exists and equals a function $K(x) \in C(\mathbb{R}^2)$ with $K \geqslant 0$ and $\int K\,dx = 1$. An analysis of the effect of this local averaging on the conserved quantities shows that the energy is retained to a good approximation, while all the nonlinear generalized enstrophies are partially lost, even in the limit as $n \to \infty$. Of course, this effect is to be expected since the ideal dynamics produce fluctuations in the vorticity on arbitrarily small scales as time proceeds, and so some information in the continuum field ω is eventually lost in the averaging that defines the observed microstate ζ. For the same reason, the prior distribution in the Miller–Robert model, which is based on exact vorticity rearrangement, is also lost. Consequently, the model proposed by Turkington relaxes the family of enstrophy constraints and chooses a different prior distribution. The precise formulation of this model is as follows.

The prior distribution ρ in the product measure (2.7) for this model is the uniform measure

$$\rho(dy) = \frac{1}{|\mathscr{Y}|}\, 1_{\mathscr{Y}}(y)\, dy \tag{2.18}$$

on the smallest interval $\mathscr{Y} = [\min \omega^0, \max \omega^0]$ containing the range of the initial state ω^0. In contrast to the choice of ρ in the Miller–Robert model, the prior distribution in (2.18) is not necessarily supported on the range of ω^0. Indeed, an observed microstate $\zeta = K_n * \omega$ will in general take values outside that range, but inside the convex hull of that range, namely, the interval $\mathscr{Y}$. The choice (2.18) has two different justifications. From an information theoretic standpoint, it is the most random, or most diffuse, measure consistent with the prior information on the distribution of ζ; namely, the lower and upper bounds $\min \omega^0 \leqslant \zeta \leqslant \max \omega^0$. From a dynamical standpoint, it defines a product measure Π_n on the phase space Ω_n that is invariant under the spectrally truncated dynamics for the lattice. That is, if $\omega(x, t)$ is replaced by a finite Fourier series defined via the discrete Fourier transform on the lattice $\mathscr{L}$ and if that Fourier series evolves under the projected equations of motion, then the induced phase flow on the microstates $\zeta = K_n * \omega \in \Omega_n$ satisfies the Liouville property. For these two reasons, the prior distribution (2.18) is the most natural choice to close the model.

The generalized enstrophy constraints for the model are derived from the fact that

$$A_n(\zeta) \doteq \frac{1}{n} \sum_{s \in \mathscr{L}} a(\zeta(s)) \leqslant A(\omega^0)$$
$$\doteq \int_{\mathscr{X}} a(\omega^0(x))\, dx \qquad \text{for all convex} \quad a \in C(\mathscr{Y})$$

This family of inequalities, which relies on Jensen's inequality, captures the effect of the local averaging $\zeta = K_n * \omega$ on any convex function a. Only the partial information contained in these convex inequalities is retained in the model, and all other generalized enstrophy constraints are considered lost. Rather than use the constraints in this form, the model replaces this family of inequalities by the equivalent family of inequalities

$$\Gamma_n(\zeta, \sigma) \doteq \frac{1}{n} \sum_{s \in \mathscr{L}} \gamma(\zeta(s), \sigma) \leqslant \Gamma(\omega^0, \sigma)$$
$$\doteq \int_{\mathscr{X}} \gamma(\omega^0(x), \sigma)\, dx \qquad \text{for all} \quad \sigma \in \mathscr{Y} \tag{2.19}$$

where $\gamma(y, \sigma)$ denotes the Green function on the interval $\mathscr{Y}$ defined by the boundary-value problem

$$\frac{d^2}{dy^2}\, \gamma(y, \sigma) = \delta(y - \sigma) \quad \text{for } y \in \mathscr{Y}, \qquad \gamma(y, \sigma) = 0 \quad \text{for } y \in \partial\mathscr{Y}$$

For each $\sigma \in \mathscr{Y}$, $\gamma(y, \sigma)$ is a piecewise linear, convex function of $y \in \mathscr{Y}$. We denote by $\mathscr{M}(\mathscr{Y})$ the space of Borel measures on $\mathscr{Y}$. Any convex function $a \in C(\mathscr{Y})$ can be represented in terms of $\gamma(y, \sigma)$ by

$$a(y) = \int_{\mathscr{Y}} \gamma(y, \sigma)\, \alpha(d\sigma) + c_0 + c_1 y \tag{2.20}$$

for some $\alpha \in \mathscr{M}(\mathscr{Y})$ and some real constants c_0 and c_1. Consequently, the lattice Gibbs state for Turkington's model is parametrized by $\beta \in \mathbb{R}$ and $\alpha \in \mathscr{M}(\mathscr{Y})$, and these parameters are dual to the constraints

$$\langle H_n(\zeta) \rangle = H(\omega^0), \qquad \langle \Gamma_n(\zeta, \sigma) \rangle \leqslant \Gamma(\omega^0, \sigma) \qquad \text{for all} \quad \sigma \in \mathscr{Y} \tag{2.21}$$

where $\langle \cdot \rangle$ denotes expectation with respect to $P_{n, \beta, a}$. This σ-parametrized family of inequalities plays the same role in this model that the global vorticity distribution constraint (2.14) plays in the Miller–Robert model.

The lattice Gibbs state for Turkington's model is most easily characterized as the solution to the minimization problem

minimize $R(P \mid \Pi_n)$ over $P \in \mathscr{P}(\Omega_n)$ subject to

$$\int_{\Omega_n} H(\zeta)\, P(d\zeta) = H(\omega^0), \qquad \int_{\Omega_n} \Gamma_n(\zeta, \sigma)\, P(d\zeta) \leqslant \Gamma(\omega^0, \sigma) \qquad \text{for all} \quad \sigma \in \mathscr{Y} \tag{2.22}$$

While the relative entropy $R(P \mid \Pi_n)$ has the same form (2.17) as in the Miller–Robert model, the product measure Π_n is defined by a different prior distribution ρ. This characterization shows that the statistical equilibrium measure $P = P_{n, \beta, a}$ defining this lattice model exists and is unique for any initial state ω^0. Indeed, by virtue of its strict convexity and lower semicontinuity, the relative entropy attains its minimum over any closed, convex set of probability measures P at a unique measure; in particular, this is the case for the closed, convex set specified by the constraints in (2.22). The solution $P_{n, \beta, a}$ to (2.22) has the canonical form (2.12), in which β is the Lagrange multiplier for the energy constraint and α is the Kuhn–Tucker vector associated with the convex enstrophy inequalities.[19, 42]

We note that one further constraint must be appended to (2.22); namely, the equality constraint on the total circulation

$$\left\langle \frac{1}{n} \sum_{s \in \mathscr{L}} \zeta(s) \right\rangle = \int_{\mathscr{X}} \omega^0(x)\, dx = 0$$

which necessarily vanishes in our prototype geometry. This constraint corresponds to the coefficient c_1 of the linear term in the representation of the arbitrary convex function a exhibited in (2.20). Since this extra term in the Gibbs measure $P_{n, \beta, a}$ can be absorbed into the function a, we shall ignore the total circulation constraint in (2.22). The modifications needed to include it are straightforward.

3. LARGE DEVIATION PRINCIPLE AND CONTINUUM LIMIT

Our main results concern the continuum limit of the lattice models introduced in the previous section. As in many other statistical mechanics problems (ref. 15, Sections 9–12), the powerful apparatus of large deviation theory can be applied to study these asymptotics. To put our problem in this general context, we need to represent the functionals that occur in the lattice Gibbs state in terms of a so-called "hidden process" on a "hidden space." In effect, the hidden process provides a macroscopic representation of the random microstate; the hidden space, in which the hidden process takes values, consists of all the admissible macrostates of the model. Specifically, we shall construct a measure-valued process Y_n whose argument is the random microstate vorticity $\zeta \in \Omega_n$ and then study the large deviation behavior of the $P_{n, n\beta, na}$-distributions of Y_n as $n \to \infty$. We shall take the hidden space to be the space of measures on $\mathscr{X} \times \mathscr{Y}$ whose first marginal is Lebesgue measure on $\mathscr{X}$.

A crucial ingredient in the continuum limit is the scaling of β and a with $n = 2^{2m}$. This particular scaling ensures that the expectations of H_n and A_n with respect to the lattice Gibbs measure remain finite as n tends to infinity; any other scaling of the parameters β and a leads to expectations that either go to zero or to infinity in the continuum limit. In essence, this scaling reflects the fact that the continuum limit is not the usual thermodynamic limit. Instead, the continuum models that we study reside on fixed domains and are defined at fixed values of total energy and generalized enstrophy, which do not scale with the number of degrees of freedom n. These fundamental features of the models are common to all statistical equilibrium theories of coherent vortex structures, whether they are built from a gas of point vortices,[17] a spectral truncation of the vorticity field,[23] or a spatial discretization of the vorticity field.[28] From the point of view of our present analysis, these known results motivate the choice of scaling parameters in the Gibbs measures $P_{n, n\beta, na}$ and hence the norming constants for the large deviation principle.

The necessary rescaling can also be accomplished by replacing the relative entropy functional in the constrained minimization problems (2.16) and (2.22), governing respectively the Miller–Robert lattice model and Turkington's lattice model, by the specific relative entropy per lattice site, $n^{-1} R(P \mid \Pi_n)$. The Lagrange multipliers for these rescaled optimization problems, say $\tilde{\beta}_n$ and $\tilde{a}_n$, then have finite limits as $n \to \infty$. Moreover, the specific relative entropy per lattice site has a finite continuum limit, which coincides with the objective functional in the maximum entropy principles for the equilibrium macrostates. An analysis along these lines is outlined in ref. 39.

In the present paper we analyze the statistical equilibrium measures in a manner that allows us to treat both the Miller–Robert model and Turkington's model simultaneously. Namely, we assume that a prior distribution ρ supported on a compact set $\mathscr{Y}$ is given together with specified parameters β and a, and we consider the lattice models defined by the scaled Gibbs measures $P_{n, n\beta, na}$. While the two models differ in how ρ, β, a are chosen and in how these quantities are determined from a given initial vorticity field, their governing Gibbs distributions have the same general form, and so our main Theorem 3.1 encompasses them both.

3.1. Statement and Proof of the Main Theorem

The space of probability measures on $\mathscr{X} \times \mathscr{Y}$, which we denote by $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$, is central to our analysis. When metrized with the metric d defined in (4.2), $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ is a Polish space (a complete, separable metric space), and convergence of measures with respect to d is equivalent to weak convergence. By Prohorov's Theorem, $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ is compact since both $\mathscr{X}$ and $\mathscr{Y}$ are compact. The large deviation principle given in Theorem 3.1 involves the relative entropy $R(\mu \mid \nu)$ of μ with respect to ν, where μ and ν are measures in $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$. This quantity is defined by

$$R(\mu \mid \nu) \doteq \begin{cases} \int_{\mathscr{X} \times \mathscr{Y}} \left(\log \dfrac{d\mu}{d\nu} \right) d\mu & \text{if} \quad \mu \ll \nu \\ \infty & \text{otherwise} \end{cases}$$

In particular, we need the relative entropy $R(\mu \mid \theta \times \rho)$ of μ with respect to the product measure $\theta \times \rho$, where $\theta(dx) = dx$ is Lebesgue measure on $\mathscr{X} = T^2$ and $\rho(dy)$ is the prior distribution on the compact subset $\mathscr{Y} \subset \mathbb{R}$. Then $R(\cdot \mid \theta \times \rho)$ defines a rate function on $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ in the sense of large deviation theory since it is a lower semicontinuous mapping of the compact space $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ into $[0, \infty]$. These standard properties are shown in Lemmas 1.4.1 and 1.4.3(b) of ref. 13.

We now list the key elements in our analysis of the continuum limit of the lattice Gibbs states by means of large deviation theory.

(i) Hidden space. This is the space $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ of probability measures on $\mathscr{X} \times \mathscr{Y}$ with first marginal θ. $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ is a closed subset of $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ and thus is also a compact Polish space.

(ii) Hidden process. This is the measure-valued random variable Y_n that maps $\Omega_n \doteq \mathscr{Y}^n$ into $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ and is defined by

$$Y_n(dx \times dy) = Y_n(\zeta, dx \times dy) \doteq \theta(dx) \otimes \sum_{s \in \mathscr{L}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy) \tag{3.1}$$

where $M(s)$ is the unique microcell of the lattice $\mathscr{L}$ containing the site s.

(iii) Energy representation function. This is the function $\tilde{H}$ that maps $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$ and is defined by

$$\tilde{H}(\mu) \doteq \tfrac{1}{2} \int_{(\mathscr{X} \times \mathscr{Y})^2} g(x - x')\, yy'\, \mu(dx \times dy)\, \mu(dx' \times dy') \tag{3.2}$$

Let $K \doteq \max\{|y|: y \in \mathscr{Y}\}$. In Lemma 4.4 we will verify that $\tilde{H}$ is bounded and continuous and that there exists $C < \infty$ such that for each n

$$\sup_{\zeta \in \Omega_n} |\tilde{H}(Y_n(\zeta)) - H_n(\zeta)| \leqslant CK^2 \left(\frac{\log n}{n}\right)^{1/2} \tag{3.3}$$

(iv) Generalized enstrophy representation function. This is the function $\tilde{A}$ that maps $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$ and is defined by

$$\tilde{A}(\mu) \doteq \int_{\mathscr{X} \times \mathscr{Y}} a(y)\, \mu(dx \times dy) \tag{3.4}$$

where a is a continuous real-valued function on $\mathscr{Y}$; the dependence of $\tilde{A}$ on a is not indicated in the notation. In Lemma 4.4 we will verify that $\tilde{A}$ is bounded and continuous and that for all $\zeta \in \Omega_n$

$$\tilde{A}(Y_n(\zeta)) = A_n(\zeta) \tag{3.5}$$

(v) Large deviation principle for Y_n with respect to Π_n. In Lemma 4.3 we will verify that Y_n satisfies the Laplace principle on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function $R(\cdot \mid \theta \times \rho)$. That is, for any bounded continuous function Φ mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$

$$\lim_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(Y_n)]\, d\Pi_n = \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{\Phi(\mu) - R(\mu \mid \theta \times \rho)\}$$

This Laplace principle is equivalent to the large deviation principle with the same rate function (ref. 13, Thms. 1.2.1 and 1.2.3); in other words, for any closed subset F of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pi_n\{Y_n \in F\} \leqslant -R(F \mid \theta \times \rho)$$

and for any open subset G of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

$$\liminf_{n \to \infty} \frac{1}{n} \log \Pi_n\{Y_n \in G\} \geqslant -R(G \mid \theta \times \rho)$$

Here, $R(B \mid \theta \times \rho)$ denotes the infimum of $R(\cdot \mid \theta \times \rho)$ over the set B.

Properties (iii)–(v) are proved in Section 4. With (i)–(v) in hand, we are able to do the following: (a) establish the asymptotic behavior of the scaled partition function $Z(n, n\beta, na)$, (b) characterize the limiting behavior of the distributions of Y_n with respect to the scaled lattice Gibbs states $P_{n, n\beta, na}$, and (c) describe the set of equilibrium macrostates for the continuum limit of the lattice model. This is the content of the following theorem, which is our main result. The physical implications of the theorem are discussed in Subsection 3.2.

Theorem 3.1. For each $\beta \in \mathbb{R}$ and $a \in C(\mathscr{Y})$, the asymptotic behavior of the model with the scaled lattice Gibbs states $P_{n, n\beta, na}$ defined in (2.12) can be described as follows.

(a)

$$\lim_{n\to\infty} \frac{1}{n} \log Z(n, n\beta, na) = -\inf_{\mu \in \mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{R(\mu \mid \theta\times\rho) + \beta\tilde{H}(\mu) + \tilde{A}(\mu)\}$$

(b) With respect to $P_{n, n\beta, na}$, the sequence Y_n satisfies the Laplace principle and the large deviation principle on $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ with rate function

$$J_{\beta, a}(\mu) \doteq R(\mu \mid \theta\times\rho) + \beta\tilde{H}(\mu) + \tilde{A}(\mu) - \inf_{\nu \in \mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{R(\nu \mid \theta\times\rho) + \beta\tilde{H}(\nu) + \tilde{A}(\nu)\}$$

That is, for any bounded continuous function Φ mapping $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ into $\mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(Y_n)]\, dP_{n, n\beta, na} = \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{\Phi(\mu) - J_{\beta, a}(\mu)\}$$

(c) Define $\mathscr{E}_{\beta, a} \doteq \{\mu \in \mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y}) : J_{\beta, a}(\mu) = 0\}$. Then $\mathscr{E}_{\beta, a}$ is a nonempty compact subset of $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$, and if B is a Borel subset of $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ whose closure $\bar{B}$ has empty intersection with $\mathscr{E}_{\beta, a}$, then $J_{\beta, a}(\bar{B}) \doteq \inf_{\mu\in\bar{B}} J_{\beta, a}(\mu) > 0$ and for some $C < \infty$

$$P_{n, n\beta, na}\{Y_n \in B\} \leqslant C \exp[-a_n J_{\beta, a}(\bar{B})] \to 0 \qquad \text{as} \quad n\to\infty$$

Remark 3.2. (a) The concentration property given in part (c) justifies calling $\mathscr{E}_{\beta, a}$ the set of equilibrium macrostates. There is an equivalent characterization of this set: by definition $\mu^* \in \mathscr{E}_{\beta, a}$ if and only if $J_{\beta, a}(\mu^*) = 0$, and this occurs if and only if μ^* gives the infimum in the variational formula in part (a). This large deviation characterization of the equilibrium macrostates is a central feature of both our approach and the approach of Michel and Robert.[27]

(b) The set $\mathscr{E}_{\beta, a}$ also arises in the study of weak limits of subsequences of the $P_{n, n\beta, a}$-distributions of Y_n, a property that further justifies calling $\mathscr{E}_{\beta, a}$ the set of equilibrium macrostates. This justification, as well as that based on part (c) of Theorem 3.1, seems clearer than the justification given in refs. 27 and 34, which is based on a concept of conditional concentration. We state the following without proof; a proof will be given in ref. 16. Any subsequence of $P_{n, n\beta, na}\{Y_n \in \cdot\}$ has a subsubsequence converging weakly to a probability measure $Q_{\beta, a}$ on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ that is concentrated on $\mathscr{E}_{\beta, a}$; i.e., $Q_{\beta, a}\{(\mathscr{E}_{\beta, a})^c\} = 0$. If $\mathscr{E}_{\beta, a}$ consists of a unique point μ, then the entire sequence $P_{n, \beta}\{Y_n \in \cdot\}$ converges weakly to δ_μ. ∎

Proof of Theorem 3.1. (a) This proof relies on the properties of $\tilde{H}$ and $\tilde{A}$ given in Lemma 4.4. By (3.3) and (3.5)

$$\left|\frac{1}{n} \log Z(n, n\beta, na) - \frac{1}{n} \log \int_{\Omega_n} \exp[-n(\beta \tilde{H}(Y_n) + \tilde{A}(Y_n))] \, d\Pi_n\right|$$

$$= \left|\frac{1}{n} \log \int_{\Omega_n} \exp[-n(\beta H_n + A_n)] \, d\Pi_n \right.$$

$$\left. - \frac{1}{n} \log \int_{\Omega_n} \exp[-n(\beta \tilde{H}(Y_n) + \tilde{A}(Y_n))] \, d\Pi_n\right|$$

$$\leqslant |\beta| \sup_{\zeta \in \Omega_n} |H_n(\zeta) - \tilde{H}(Y_n(\zeta))| \leqslant CK^2 \, |\beta| \left(\frac{\log n}{n}\right)^{1/2}$$

Since $\tilde{H}$ and $\tilde{A}$ are bounded continuous functions mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$ [Lemma 4.4(a), (c)], the Laplace Principle satisfied by Y_n with respect to Π_n [Lemma 4.3] yields

$$\lim_{n \to \infty} \frac{1}{n} \log Z(n, n\beta, na)$$

$$= \lim_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[-n(\beta \tilde{H}(Y_n) + \tilde{A}(Y_n))] \, d\Pi_n$$

$$= \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{-\beta \tilde{H}(\mu) - \tilde{A}(\mu) - R(\mu \mid \theta \otimes \rho)\}$$

This proves part (a).

(b) The proof of this part uses the same methods as part (a), but now with $P_{n, n\beta, na}$ replacing Π_n. For any bounded continuous function Φ mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$, again (3.3) and (3.5) yield

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(Y_n)]\, dP_{n,n\beta,na}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n(\Phi(Y_n) - \beta H_n - A_n)]\, d\Pi_n$$

$$- \lim_{n\to\infty} \frac{1}{n} \log Z(n, n\beta, na)$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n(\Phi(Y_n) - \beta\tilde{H}(Y_n) - \tilde{A}(Y_n))]\, d\Pi_n$$

$$- \lim_{n\to\infty} \frac{1}{n} \log Z(n, n\beta, na)$$

$$= \sup_{\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{\Phi(\mu) - \beta\tilde{H}(\mu) - \tilde{A}(\mu) - R(\mu \mid \theta\times\rho)\}$$

$$+ \inf_{\nu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{\beta\tilde{H}(\nu) + \tilde{A}(\nu) + R(\nu \mid \theta\times\rho)\}$$

$$= \sup_{\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{\Phi(\mu) - J_{\beta,a}(\mu)\}$$

$J_{\beta,a}$ is a lower semicontinuous mapping of the compact space $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ into $[0,\infty]$ and so defines a rate function. We conclude that with respect to $P_{n,n\beta,na}$ the sequence Y_n satisfies the Laplace principle on $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ with rate function $J_{\beta,a}$. Since this is equivalent to the large deviation principle with the same rate function, part (b) is proved.

(c) Since $J_{\beta,a}$ is a rate function, it assumes its minimum of 0 on $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$. This gives the first assertion. The second assertion is a consequence of the fact that if $\bar{B}$ has empty intersection with $\mathscr{E}_{\beta,a}$, then $J_{\beta,a}(\mu) > 0$ for each $\mu\in\bar{B}$. Since $J_{\beta,a}$ is a rate function, $J_{\beta,a}(\bar{B}) > 0$. The large deviation upper bound proved in part (b) completes the proof of part (c). ∎

3.2. Interpretation of the Main Results

We now connect the mathematical results in Theorem 3.1 to the statistical equilibrium models of coherent states in ideal two-dimensional turbulence, thereby elucidating the physical implications of these results.

The hidden space $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ consists of all possible macrostates $\mu = \mu(dx\times dy)$ for the continuum models. In order to see how these measures offer a macroscopic description of the vorticity field, we decompose any such $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ as

$$\mu(dx \times dy) = \theta(dx) \otimes \tau(x, dy)$$

where $\theta(dx) = dx$ on $\mathscr{X}$ and $\tau(x, dy)$ is a stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$. That is, $\tau(x, dy)$ is a family of probability measures on $\mathscr{Y}$ indexed by $x \in \mathscr{X}$ such that for each Borel subset C of $\mathscr{Y}$ the mapping $x \mapsto \tau(x, C)$ is measurable, and for each Borel subset B of $\mathscr{X} \times \mathscr{Y}$

$$\mu\{B\} = \int_B \theta(dx)\, \tau(x, dy)$$

We summarize this decomposition by the notation $\mu = \theta \otimes \tau$. A proof is given in (ref. 13, Thm. A.5.4).

We recall that for $\beta \in \mathbb{R}$ and $a \in C(\mathscr{Y})$ the set $\mathscr{E}_{\beta, a}$ of equilibrium macrostates is defined by

$$\mathscr{E}_{\beta, a} \doteq \{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}) : J_{\beta, a}(\mu) = 0\}$$

where

$$J_{\beta, a}(\mu) \doteq R(\mu \mid \theta \times \rho) + \beta \tilde{H}(\mu) + \tilde{A}(\mu) - \inf_{\nu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{R(\nu \mid \theta \times \rho) + \beta \tilde{H}(\nu) + \tilde{A}(\nu)\}$$

In particular, if $\mu \in \mathscr{E}_{\beta, a}$, then $R(\mu \mid \theta \times \rho) < \infty$; thus $\mu \ll \theta \times \rho$, and so the stochastic kernel $\tau(x, dy)$ appearing in the decomposition $\mu = \theta \otimes \tau$ satisfies $\tau(x, \cdot) \ll \rho$ for almost all $x \in \mathscr{X}$. For equilibrium macrostates μ, this property allows us to refine the decomposition of μ. Indeed, for almost all $x \in \mathscr{X}$ there exists a density function $p(x, \cdot) \doteq d\tau(x, \cdot)/d\rho$ such that $p(x, \cdot) \in L^1(\mathscr{Y})$ and

$$\mu(dx \times dy) = \theta(dx) \otimes [\, p(x, y)\, \rho(dy)]$$

This refined decomposition shows that an equilibrium macrostate μ is uniquely specified by an x-parametrized family of probability densities on $\mathscr{Y}$ with respect to the prior distribution ρ. We can now make the following intuitive interpretation of such a macrostate: $p(x, y)\, \rho(dy)$ is the probability that the random vorticity field takes a value in an infinitesimal interval dy when sampled in an infinitesimally small neigborhood dx of x. A macroscopic description of this kind is necessary because the underlying ideal fluid dynamics can excite fluctuations of vorticity on arbitrarily fine scales at any point of the domain; it is sufficient because the correlation between fluctuations at any two distinct points x and x' vanishes in the continuum limit. The use of these macrostates is an innovation of ref. 34, which calls them Young measures by analogy with the parametrized measures that arise in the study of weak limits of solutions to nonlinear partial differential equations. However, these macrostates in the statistical equilibrium theory have a natural probabilistic meaning as the one-point distributions of the random vorticity. They capture the long-time average behavior of the microscopic vorticity, not merely a possible weak limit of solutions $\omega(\cdot, t)$ of (2.1) over some sequence of times $t = t_j \to \infty$.

The hidden process Y_n establishes the link between the microstate $\zeta \in \Omega_n$, which provides the fine-grained description of the vorticity field, and the macrostate $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, which furnishes the corresponding coarse-grained description. The central result in Theorem 3.1 is the large deviation principle for Y_n with respect to the scaled lattice Gibbs measures $P_{n, n\beta, na}$. This result reveals that Y_n concentrates on the set $\mathscr{E}_{\beta, a}$ of equilibrium macrostates, which minimize $R(\cdot \mid \theta \times \rho) + \beta \tilde{H} + \tilde{A}$.

From a technical viewpoint, we choose the hidden process Y_n to reduce the large deviation analysis of the scaled lattice Gibbs states to a straightforward application of Laplace's principle. To this purpose we defined Y_n so that, first, it yields representations $\tilde{H}$ and $\tilde{A}$ of the functions H_n and A_n determining the Gibbs weight, either as a suitable approximation or an exact expression and, second, it satisfies a large deviation principle with respect to the product measures Π_n built from the prior distribution ρ. The first requirement is met by constructing Y_n from δ-measures on the range of vorticity $\mathscr{Y}$ spread over the domain $\mathscr{X}$. The second requirement, that Y_n satisfy a large deviation principle with respect to Π_n, is more subtle because this depends crucially on the topology of the hidden space $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. Heuristically, we can understand this large deviation principle by saying that, for any given macrostate $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, there are many microstates ζ for which $Y_n(\zeta, \cdot)$ is close to μ with respect to a suitable metric d to be defined in Section 4.1; the multiplicity of these microstates is quantified by the relative entropy $R(\mu \mid \theta \times \rho)$.

In Section 4, instead of proving this rather subtle large deviation principle directly, we offer a more intuitive approach. There, we approximate Y_n by a doubly indexed process $W_{n,r}$ consisting of 2^r empirical measures, each constructed by aggregating microcells locally in the lattice. By this technique, the large deviation principle for Y_n is reduced to a large deviation principle for $W_{n,r}$. This approach has the virtue that the large deviation principle for $W_{n,r}$ can be regarded as a two-parameter version of the familiar Sanov theorem for empirical measures, from which it follows quite naturally. The construction of $W_{n,r}$ is based on a local averaging over an intermediate length scale controlled by the second index r. This construction not only explains the separation-of-scales behavior of these statistical equilibrium models, but also is the mathematical expression of the physical reasoning pioneered by Miller *et al.*[28, 29]

We now turn to the physical implications of the large deviation principle proved in Theorem 3.1. For the sake of definiteness in this discussion, let us assume that the set $\mathscr{E}_{\beta, a}$ of equilibrium macrostates defined in part (c) of Theorem 3.1 consists of a unique macrostate $\mu = \mu_{\beta, a}$. This assumption allows us to ignore degeneracies when explaining the physical implications of the mathematical results.

Let f be any function in $C(\mathscr{X} \times \mathscr{Y})$. The essential content of the abstract statement that with respect to the statistical equilibrium measures $P_{n, n\beta, na}$ the sequence of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$-valued random variables Y_n satisfies the large deviation principle can be expressed in concrete terms by means of the class of observables

$$F_n(\zeta) \doteq \frac{1}{n} \sum_{s \in \mathscr{L}} f(s, \zeta(s)) \tag{3.6}$$

and their representation functions

$$\tilde{F}(\mu) \doteq \int_{\mathscr{X} \times \mathscr{Y}} f(x, y)\, \mu(dx \times dy)$$

With a suitable choice of f, these observables F_n are capable of discriminating the macroscopic properties of the vorticity field in the model. For example, fix $x^* \in \mathscr{X}$ and consider $f^\varepsilon(x, y) = \varepsilon^{-2} k((x - x^*)/\varepsilon)\, y$, where $k(x)$ is a nonconstant continuous function on $\mathscr{X}$ satisfying $k \geqslant 0$, $\int k\, dx = 1$, and having a unique maximum at 0. Then for small $\varepsilon > 0$ the corresponding observable F_n^ε captures the local average of ζ near x^*. For any observable F_n having the form (3.6), a direct consequence of the large deviation principle is the fact that for any $\eta > 0$

$$P_{n, n\beta, na}\{ |F_n - \tilde{F}(\mu_{\beta, a})| \geqslant \eta \} \to 0 \qquad \text{as} \quad n \to \infty \tag{3.7}$$

To see this, we first note that by the uniform continuity of f

$$\sup_{\zeta \in \Omega_n} |F_n(\zeta) - \tilde{F}(Y_n(\zeta))| \to 0 \qquad \text{as} \quad n \to \infty$$

For sufficiently large n depending on η, we then find that

$$\begin{aligned} P_{n, n\beta, na}\{ |F_n - \tilde{F}(\mu_{\beta, a})| \geqslant \eta \} &\leqslant P_{n, n\beta, na}\{ |\tilde{F}(Y_n) - \tilde{F}(\mu_{\beta, a})| \geqslant \eta/2 \} \\ &\leqslant \exp(-n I_\eta/2) \end{aligned}$$

where

$$I_\eta \doteq \inf\{J_{\beta,a}(\mu): \mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}), |\tilde{F}(\mu) - \tilde{F}(\mu_{\beta,a})| \geqslant \eta/2\} > 0$$

The limit (3.7) now follows. This property shows that in the continuum limit each observable F_n concentrates around the value of its representation function $\tilde{F}$ at the equilibrium macrostate $\mu_{\beta,a}$.

We now turn our attention to the first-order variational conditions for the equilibrium macrostates μ, which are characterized as zeroes of the rate function $J_{\beta,a}$ or equivalently as minimizers of $R(\cdot \mid \theta \times \rho) + \beta\tilde{H} + \tilde{A}$. These first-order variational conditions are most readily expressed in terms of the decomposition $\mu(dx \times dy) = \theta(dx) \otimes [p(x, y)\, \rho(dy)]$. As shown in refs. 34 and 39, the variational conditions reduce to the equation

$$p(x, y) = \frac{\exp(-\beta\bar{\psi}(x)\, y - a(y))}{\int_{\mathscr{Y}} \exp(-\beta\bar{\psi}(x)\, y - a(y))\, \rho(dy)} \tag{3.8}$$

where the mean streamfunction $\bar{\psi} = G\bar{\omega}$ is determined as in (2.2); $\bar{\omega}$ denotes the mean vorticity

$$\bar{\omega}(x) \doteq \int_{\mathscr{Y}} yp(x, y)\, \rho(dy)$$

Equation (3.8) is implicit and nonlinear in p because of the dependence on the mean streamfunction $\bar{\psi}$. This equation for p reduces to a nonlinear elliptic equation for $\bar{\psi}$ itself upon taking the expectation over y; namely,

$$\bar{\omega} = -\Delta\bar{\psi} = \Lambda'_{\beta,a}(\bar{\psi}) \tag{3.9}$$

where for $\varphi \in \mathbb{R}$

$$\Lambda_{\beta,a}(\varphi) \doteq -\frac{1}{\beta} \log \int_{\mathscr{Y}} \exp(-\beta\varphi y - a(y))\, \rho(dy)$$

The mean-field equation (3.9) is one of the major predictions of the theory. It ensures that the mean flow is steady, and moreover it supplies the particular functional dependence of the mean vorticity on the mean streamfunction. From a deterministic viewpoint, the vorticity-streamfunction profile $\bar{\omega} = \Lambda'(\bar{\psi})$ can be arbitrary for steady flow, although the further condition of stability requires that Λ' be monotonic. On the other hand, each statistical equilibrium model produces a distinguished profile and explains how this profile is determined by the parameters β and a.

Various special cases of the theory are examined in detail in the literature, all based upon some simple choice of prior distribution and a corresponding reduction of the continuum of enstrophy integrals. Several such reductions are detailed in ref. 39. For the purpose of the present summary, it suffices to mention the simplest case applicable to the torus geometry; namely, vortex patches on $\mathscr{X} = T^2$. The initial vorticity field in this case takes on two values; say, $\omega^0 = +1$ in $U \subset \mathscr{X}$ and $\omega^0 = -1$ in $\mathscr{X} \setminus U$, where the area of U is half the area of $\mathscr{X}$. Then it can be shown that the enstrophy constaints in both the Miller–Robert model and Turkington's model play no role, and hence the function a appearing in $\Lambda_{\beta, a}$ reduces to $a(y) = \alpha y$, where α is the multiplier for the total circulation constraint. Consequently, the mean-field equation takes an especially simple form in either model. In the Miller–Robert model it is

$$-\Delta\bar{\psi} = \tanh(-\alpha - \beta\bar{\psi})$$

while in Turkington's model it is

$$-\Delta\bar{\psi} = \coth(-\alpha - \beta\bar{\psi}) - (-\alpha - \beta\bar{\psi})^{-1}$$

In each model the ergodic mixing of the two-level initial vorticity ω^0 results in a final vorticity $\bar{\omega}$ with a distinctive profile that mediates between those extreme levels. In both models $\beta < 0$, and so the coherent states have negative temperature. The contrast between the prediction of the two models is discussed in detail in ref. 39. In brief, Turkington's model allows a mixing of microscopic vorticity on a range of small scales, while the Miller–Robert model enforces a single microscale on the mixing. The prior distribution $\rho(dy)$ in the Miller–Robert model is atomic at $y = \pm 1$, while in Turkington's model it is uniform on the interval $-1 \leqslant y \leqslant +1$. The macroscopic manifestation of the difference at the microscopic scales is that the vorticity-streamfunction profile in Turkington's model is more gradual and takes the asymptotic values ± 1 more slowly than the corresponding profile in the Miller–Robert model.

3.3. Maximum Entropy Principles

While the parametrization of the set $\mathscr{E}_{\beta, a}$ of equilibrium macrostates by the inverse temperature β and the enstrophy profile function a is convenient in the analysis of the statistical equilibrium lattice models, it is not the most natural way to parametrize the set of equilibrium macrostates in the continuum models. Instead, for applications to fluid dynamics, it is often better to parametrize the set of equilibrium macrostates by the values of the energy and the generalized enstrophy. One views these values as parameters derived from the given initial data, while one regards the corresponding β and a as unknowns that are determined along with the equilibrium macrostate. This reformulation conforms with the idea that the equilibrium macrostate describes the final mixed state that ensues after an ergodic evolution from a specified initial state. In this context it is useful to view the equilibrium macrostates in the continuum models as being governed by *constrained* minimization problems, in which the objective functional is the relative entropy and the constraints involve the energy representation function and the global vorticity distribution; the constraint values are derived from an initial vorticity field ω^0. The parameters β and a are then realized as Lagrange multipliers dual to these respective constraints. The possibility of two different parametrizations of the equilibrium macrostates—on the one hand, by the inverse temperature β and the enstrophy profile function a and on the other hand, by values of the energy and the global vorticity distribution—involves issues of the equivalence of ensembles, which we shall briefly address later in this subsection.

We now proceed to state these constrained minimization problems, considering the Miller–Robert model and Turkington's model separately. Following the usual convention in the physical literature, in both cases we refer to them as "maximum entropy principles" or as "constrained maximum entropy principles" even though we minimize the relative entropy R, which differs from the physical entropy $S \doteq -R$ by a minus sign.

By a calculation at the physical level of rigor, the paper (ref. 34) shows that in the continuum limit, equilibrium macrostates μ in the Miller–Robert model are solutions of the constrained maximum entropy principle

$$\text{minimize } R(\mu \mid \theta \times \rho) \text{ over } \mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}) \text{ subject to}$$

$$\tilde{H}(\mu) = H(\omega^0), \qquad \int_{\mathscr{X}} \mu(dx \times \cdot) = \int_{\mathscr{X}} \delta_{\omega^0(x)}(\cdot)\, dx \tag{3.10}$$

In this model the prior distribution $\rho(\cdot)$ is defined by the second integral in the preceding display and is supported on the closure of range of ω^0. The constrained maximum entropy principle (3.10) is particularly appealing since it is an obvious analogue of the constrained minimization problem (2.16) that defines the lattice Gibbs measure $P_{n,\beta,a}$. The variational calculation for solutions to (3.10) is discussed in ref. 34, and so we shall not present a rigorous justification of it here.

We now turn to Turkington's model. By a calculation at the physical level of rigor, the paper (ref. 39) shows that in the continuum limit, equilibrium macrostates μ in Turkington's model are solutions of the constrained maximum entropy principle

$$\text{minimize } R(\mu \mid \theta \times \rho) \text{ over } \mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}) \text{ subject to } \tilde{H}(\mu) = H(\omega^0),$$

$$\int_{\mathscr{X} \times \mathscr{Y}} \gamma(y, \sigma)\, \mu(dx \times dy) \leqslant \int_{\mathscr{X}} \gamma(\omega^0(x), \sigma)\, dx \qquad \text{for all} \quad \sigma \in \mathscr{Y} \tag{3.11}$$

In contrast to the Miller–Robert model, the prior distribution ρ for this model is the uniform measure (2.18) on the smallest closed interval $\mathscr{Y}$ containing the range of ω^0. Also, the Miller–Robert constraint on the global vorticity distribution is relaxed in this model to a family of convex inequalities parametrized by $\sigma \in \mathscr{Y}$. The variational calculation for solutions to (3.11) involves the Kuhn–Tucker conditions. It is discussed in ref. 39.

We now consider the relationship between the set $\mathscr{E}_{\beta, a}$ of equilibrium macrostates, which is defined in part (c) of Theorem 3.1, and solutions of the constrained maximum entropy principles (3.10) and (3.11). To simplify the discussion, we shall restrict our comments to the Miller–Robert model; Turkington's model is handled similarly. By definition, for given $\beta \in \mathbb{R}$ and $a \in C(\mathscr{Y})$ a measure μ belongs to $\mathscr{E}_{\beta, a}$ if and only if μ minimizes $R(\cdot \mid \theta \times \rho) + \beta \tilde{H} + \tilde{A}$. One can show that the first-order variational conditions for this minimization problem coincide with the first-order variational conditional conditions for the constrained maximum entropy principle (3.10), where β and a arise as Lagrange multipliers. These first-order variational conditions are given by (3.8). However, the correspondence between *minimizers* of $R(\cdot \mid \theta \times \rho) + \beta \tilde{H} + \tilde{A}$ and solutions of the maximum entropy principle is not obvious.

It is useful to investigate this relationship in the context of the equivalence of ensembles; namely, the canonical ensemble, which is defined by the scaled Gibbs measures $P_{n, n\beta, na}$, and the microcanonical ensemble, which ideally is defined by conditioning the product measures $\Pi_n(d\zeta) \doteq \prod_{s \in \mathscr{L}} \rho(d\zeta(s))$ on configurations satisfying the following constraints:

$$H_n(\zeta) = E, \qquad \frac{1}{n} \sum_{\zeta \in \mathscr{L}} \delta_{\zeta(s)}(dy) = \rho(dy) \tag{3.12}$$

These are microcanonical analogues of (2.15) and (2.14), respectively. In a sequel to the present paper,[16] we prove a conditional large deviation principle for the hidden process Y_n with respect to product measures conditioned on sets that approximate these constraints. For ease of exposition in this paragraph, we will refer to the constraints (3.12) given in their ideal form, even though the theorems in ref. 16 involve some approximation in the conditioning. We label the rate function in the conditional large deviation principle as $I^{E,\rho}$. By analogy with the definition of $\mathscr{E}_{\beta,a}$ as the set of minimizers of the rate function $J_{\beta,a}$ in Theorem 3.1, we define $\mathscr{E}^{E,\rho}$ to be the set of $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ minimizing the rate function $I^{E,\rho}$ in the conditional large deviation principle. The form of this rate function shows that μ minimizes $I^{E,\rho}$ if and only if μ solves the maximum entropy principle (3.10). In essentially the same way that we justified the designation of $\mathscr{E}_{\beta,a}$ as the set of equilibrium macrostates with respect to the $P_{n,n\beta,na}$-distributions [Theorem 3.1(c) and Remark 3.2(b)], one can justify the designation of $\mathscr{E}^{E,\rho}$ as the set of equilibrium macrostates with respect to the conditioned Π_n-distributions of Y_n. In ref. 16 the relationship between canonical equilibrium macrostates and microcanonical equilibrium macrostates is investigated under appropriate assumptions on the model.

We next show a simple but important relationship between canonical equilibrium macrostates and solutions of the constrained maximum entropy principle (3.10). Namely, for given $\beta \in \mathbb{R}$ and $a \in C(\mathscr{Y})$, any equilibrium macrostate $\mu_{\beta,a} \in \mathscr{E}_{\beta,a}$ solves the constrained maximum entropy principle (3.10) when the constraint values in this variational problem are derived from $\mu_{\beta,a}$ itself; i.e.,

$$\tilde{H}(\mu) = \tilde{H}(\mu_{\beta,a}), \qquad \int_{\mathscr{X}} \mu(dx \times \cdot\,) = \int_{\mathscr{X}} \mu_{\beta,a}(dx \times \cdot\,) \tag{3.13}$$

Since $\mu_{\beta,a} \in \mathscr{E}_{\beta,a}$, for any $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ we have

$$R(\mu_{\beta,a} \mid \theta \times \rho) + \beta \tilde{H}(\mu_{\beta,a}) + \tilde{A}(\mu_{\beta,a}) \leqslant R(\mu \mid \theta \times \rho) + \beta \tilde{H}(\mu) + \tilde{A}(\mu)$$

If μ also satisfies the constraints in (3.13), then the second constraint implies that

$$\tilde{A}(\mu) = \int_{\mathscr{X} \times \mathscr{Y}} a(y)\, \mu(dx \times dy) = \int_{\mathscr{X} \times \mathscr{Y}} a(y)\, \mu_{\beta,a}(dx \times dy) = \tilde{A}(\mu_{\beta,a})$$

Hence for all such μ, $R(\mu_{\beta,a} \mid \theta \times \rho) \leqslant R(\mu \mid \theta \times \rho)$. We conclude that $\mu_{\beta,a}$ satisfies the maximum entropy principle (3.10) with the constraints (3.13). This proves the claim.

This result and its analogue for Turkington's model demonstrate that the constrained maximum entropy principles governing these continuum models are derived from Gibbsian equilibrium statistical mechanics by means of the large deviation principle in Theorem 3.1, at least over the range of constraint values determined by equilibrium macrostates. Within the context of canonical ensembles, this result is complete in the sense that every equilibrium macrostate is identified with a solution to the corresponding constrained maximum entropy principle. However, in order to establish a one-to-one correspondence between equilibrium macrostates and these solutions, it is necessary to impose further conditions that ensure the equivalence of the canonical and microcanonical ensembles. An analysis of the microcanonical ensemble and a general sufficient condition for equivalence of ensembles is given in a subsequent paper (ref. 16).

For the purpose of the present discussion, let us illustrate this condition in the special case of the Miller–Robert model already mentioned in the last paragraph of Subsection 3.2. Namely, we take the prior distribution to be $\rho(dy) \doteq \frac{1}{2}\delta_1(dy) + \frac{1}{2}\delta_{-1}(dy)$, for which one can show that the enstrophy constraints in (3.10) reduce to the compatibility condition $\int_{\mathscr{X}} \bar{\omega}\, dx = 0$. In this case a sufficient condition that the microcanonical equilibrium set $\mathscr{E}^{E,\rho}$ coincide with the corresponding canonical equilibrium set is that the equilibrium relative entropy $R(\mathscr{E}^{E,\rho} \mid \theta \times \rho)$ be a strictly convex function of E. In turn, this convexity condition can be verified numerically by solving the constrained maximum entropy problem over the range of admissible energy values $E \in [0, E_{\max}]$. A numerical method for solving problems of this kind is presented in ref. 40. It is an iterative algorithm that increases physical entropy at every iteration, and it is provably convergent to the equilibrium set $\mathscr{E}^{E,\rho}$ from any admissible initial guess. Computation of the solution sets in this simple case of the Miller–Robert model on the unit torus shows that the required strict convexity property is indeed valid. Numerous other cases of physical interest have also been computed.[2, 5, 11, 20, 40, 41]

4. PROPERTIES OF THE HIDDEN PROCESS

The proof of Theorem 3.1 relies on properties of the hidden process Y_n, the energy representation function $\tilde{H}$, and the generalized enstrophy representation function $\tilde{A}$, which are defined in (3.1), (3.2), and (3.4), respectively. In this section we shall complete the proof of our main theorem by establishing these properties.

4.1. Asymptotics of the Hidden Process

A key step in the proof of Theorem 3.1 is to show the Laplace principle for the hidden process

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

with respect to the product measures Π_n. One of the key innovations of this paper is to prove the Laplace principle for Y_n by approximating Y_n by another, doubly indexed process for which the Laplace principle can be shown directly. This doubly indexed process is built from local averages of Y_n and is defined as follows. Given $n = 2^{2m}$, we consider a dyadic partition of the lattice $\mathscr{L}$ into 2^r blocks, with r even and $r < 2m$, each block containing $n/2^r$ lattice sites. In correspondence with this partition of the lattice into blocks, we have a dyadic partition $\{D_{r,k}, k = 1,..., 2^r\}$ of the torus $\mathscr{X} = T^2$ into macrocells of the form

$$[(i-1)/2^{r/2}, i/2^{r/2}) \times [(j-1)/2^{r/2}, j/2^{r/2}), \qquad i, j \in \{1,..., 2^{r/2}\}$$

Each such macrocell $D_{r,k}$ is the union of $n/2^r$ microcells $M(s)$, each of which is a square having area $1/n$ and containing one site s of $\mathscr{L}$. With respect to a partition of this kind, we define for $\zeta \in \Omega_n \doteq \mathscr{Y}^n$

$$W_{n,r}(dx \times dy) = W_{n,r}(\zeta, dx \times dy) \doteq dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x) \frac{1}{n/2^r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(dy) \tag{4.1}$$

$W_{n,r}$ is obtained from Y_n by replacing, for each $s \in D_{r,k}$, the point mass $\delta_{\zeta(s)}$ by the average $(n/2^r)^{-1} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}$ over the $n/2^r$ sites contained in $D_{r,k}$. $W_{n,r}$ is a random measure taking values in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$.

Our purpose in introducing the doubly indexed process $W_{n,r}$ can now be explained. The spatial scale of the sets $D_{r,k}$ is intermediate between the scale of the torus $\mathscr{X}$ and the scale of the microcells $M(s)$. The local averaging over the sets $D_{r,k}$ produces $W_{n,r}$, which is a sum of terms

$$L_{n,r,k}(dy) = L_{n,r,k}(\zeta, dy) \doteq \frac{1}{n/2^r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(dy)$$

indexed by $k = 1,..., 2^r$. Since each of these terms is an empirical measure for the i.i.d. sequence $\{\zeta(s), s \in D_{r,k}\}$, for each r and $k \in \{1,..., 2^r\}$ Sanov's Theorem immediately implies that $L_{n,r,k}$ satisfies a large deviation principle as n tends to infinity.[9, 10, 13] As we will see, it then follows that the doubly indexed process $W_{n,r}$ satisfies a two-parameter large deviation principle as n and r go to infinity. This key result is stated in Lemma 4.1 in the form of an equivalent Laplace principle. Then in Lemma 4.2 we show that as r tends to infinity the distance between Y_n and $W_{n,r}$ in an appropriate metric goes to zero uniformly in n. From this appoximation we readily obtain the desired Laplace principle for Y_n, which is stated in Lemma 4.3.

Lemma 4.1. With respect to the measures Π_n, the sequence $W_{n,r}$ satisfies the two-parameter Laplace principle on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function $R(\cdot \mid \theta \times \rho)$. In other words, for any bounded continuous function Φ mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$

$$\lim_{r \to \infty} \lim_{n \to \infty} \frac{1}{n} \log \int_{\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \exp[n\Phi(W_{n,r})] \, d\Pi_n$$

$$= \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{\Phi(\mu) - R(\mu \mid \theta \times \rho)\}$$

This lemma is a special case of a general theorem (ref. 1, Thm. 4.1) that applies to a class of processes that includes the random measures $W_{n,r}$. Because of the centrality of Lemma 4.1 in the present paper, we prove it in Section 5. There it is stated in the equivalent form of a two-parameter large deviation principle for $W_{n,r}$ on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function $R(\cdot \mid \theta \times \rho)$. The proof of the equivalence of this large deviation principle with the Laplace principle carries over by making obvious modifications in the proof for singly indexed processes (ref. 13, Thms. 1.2.1 and 1.2.3).

The approximation result relating Y_n and $W_{n,r}$ uses a metric on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ that is compatible with the topology of weak convergence of measures. Let $BL(\mathscr{X} \times \mathscr{Y})$ denote the space of bounded, Lipschitz continuous functions f mapping $\mathscr{X} \times \mathscr{Y}$ into $\mathbb{R}$ and define a norm on this space by

$$\|f\|_{BL} \doteq \sup_{(x,y)} |f(x,y)| + \sup_{(x,y) \neq (x',y')} \frac{|f(x,y) - f(x',y')|}{|x - x'| + |y - y'|}$$

Then metrize $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ with the dual-bounded-Lipschitz metric

$$d(\mu, \nu) \doteq \sup \left\{ \left| \int_{\mathscr{X} \times \mathscr{Y}} f \, d\mu - \int_{\mathscr{X} \times \mathscr{Y}} f \, d\nu \right| : f \in BL(\mathscr{X} \times \mathscr{Y}) \text{ with } \|f\|_{BL} \leqslant 1 \right\} \tag{4.2}$$

When metrized by d, the hidden space $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, being a closed subspace of $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$, is a compact Polish space (ref. 12, Prop. 11.3.2, Thm. 11.3.3, Cor. 11.5.5). The approximation result is stated next.

Lemma 4.2. For all $n = 2^{2m}$ and even $r \in \mathbb{N}$ satisfying $r < 2m$, $d(Y_n, W_{n,r}) \leqslant \sqrt{2}/2^{r/2}$.

Proof. We first rewrite Y_n and $W_{n,r}$ in a common form by introducing, for each $k \in \{1,..., 2^r\}$, a sum over the $n/2^r$ sites in $D_{r,k} = \bigcup_{s' \in D_{r,k}} M(s')$. Namely,

$$Y_n(dx \times dy) = dx \otimes \sum_{k=1}^{2^r} \sum_{s \in D_{r,k}} \frac{1}{n/2^r} \sum_{s' \in D_{r,k}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy)$$

$$W_{n,r}(dx \times dy) = dx \otimes \sum_{k=1}^{2^r} \sum_{s \in D_{r,k}} \frac{1}{n/2^r} \sum_{s' \in D_{r,k}} 1_{M(s')}(x)\, \delta_{\zeta(s)}(dy)$$

Then for any $f \in BL(\mathscr{X} \times \mathscr{Y})$ with $\|f\|_{BL} \leqslant 1$, we have

$$\begin{aligned}
&\left| \int_{\mathscr{X} \times \mathscr{Y}} f\, dY_n - \int_{\mathscr{X} \times \mathscr{Y}} f\, dW_{n,r} \right| \\
&\quad = \frac{1}{n/2^r} \left| \sum_{k=1}^{2^r} \sum_{s, s' \in D_{r,k}} \left(\int_{M(s)} f(x, \zeta(s))\, dx - \int_{M(s')} f(x', \zeta(s))\, dx' \right) \right| \\
&\quad = \frac{n}{n/2^r} \left| \sum_{k=1}^{2^r} \sum_{s, s' \in D_{r,k}} \int_{M(s)} \int_{M(s')} [f(x, \zeta(s)) - f(x', \zeta(s))]\, dx\, dx' \right| \\
&\quad \leqslant \frac{n}{n/2^r} \sum_{k=1}^{2^r} \sum_{s, s' \in D_{r,k}} \int_{M(s)} \int_{M(s')} |x - x'|\, dx\, dx' \\
&\quad = 2^r \sum_{k=1}^{2^r} \int_{D_{r,k}} \int_{D_{r,k}} |x - x'|\, dx\, dx' \\
&\quad \leqslant \max_{k=1,..., 2^r} \operatorname{diam}(D_{r,k}) = \sqrt{2}/2^{r/2}
\end{aligned}$$

It follows that $d(Y_n, W_{n,r}) \leqslant \sqrt{2}/2^{r/2}$, as claimed. ∎

Finally, we derive the desired Laplace principle for Y_n from the two-parameter Laplace principle for $W_{n,r}$.

Lemma 4.3. With respect to the measures Π_n, the sequence of random measures Y_n defined in (3.1) satisfies the Laplace principle on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function $R(\cdot \mid \theta \times \rho)$. In other words, for any bounded continuous function Φ mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$

$$\lim_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(Y_n)]\, d\Pi_n = \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{\Phi(\mu) - R(\mu \mid \theta \times \rho)\} \tag{4.3}$$

Proof. By [ref. 13, Cor. 1.2.5], it suffices to prove the Laplace limit (4.3) for all bounded, Lipschitz continuous functions Φ mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

into $\mathbb{R}$. Let Φ be such a function with Lipschitz constant M and take $\delta > 0$. The previous lemma and the two-parameter Laplace principle for $W_{n,r}$ in Lemma 4.1 imply the following: for all sufficiently large, even r, we have $d(Y_n, W_{n,r}) \leqslant \delta$ for all $n = 2^{2m}$ satisfying $2m > r$ and

$$\limsup_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(W_{n,r})] \, d\Pi_n \leqslant \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{\Phi(\mu) - R(\mu \mid \theta \times \rho)\} + \delta$$

For all such r

$$\begin{aligned}\limsup_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(Y_n)] \, d\Pi_n &\leqslant \limsup_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(W_{n,r}) + nMd(Y_n, W_{n,r})] \, d\Pi_n \\ &\leqslant \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{\Phi(\mu) - R(\mu \mid \theta \times \rho)\} + \delta + M\delta\end{aligned}$$

Similarly,

$$\liminf_{n \to \infty} \frac{1}{n} \log \int_{\Omega_n} \exp[n\Phi(Y_n)] \, d\Pi_n \geqslant \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})} \{\Phi(\mu) - R(\mu \mid \theta \times \rho)\} - \delta - M\delta$$

Since $\delta > 0$ is arbitrary, the proof of the lemma is complete. ▮

4.2. Properties of $\tilde{H}$ and $\tilde{A}$

It remains to show the properties of $\tilde{H}$ and $\tilde{A}$ asserted in (iii) and (iv) at the beginning of Subsection 3.1. The identity (3.5) involving $\tilde{A}$ and A_n is immediate from the definition of Y_n. On the other hand, because of the singularity of the Green's function $g(x - x')$ along the diagonal $x = x'$, some effort is required to show that $\tilde{H}$ evaluated at the hidden process Y_n provides the approximation to the Hamiltonian H_n specified in (3.3). For this proof we use the Fourier representations of g and of the lattice Green's functions g_n, thereby exploiting the prototype geometry $\mathscr{X} = T^2$. A different proof that would apply to other geometries and boundary conditions could be based on the mildness of the singularity in g, but it would be more complicated in our case.

Lemma 4.4. The energy representation function $\tilde{H}$ defined in (3.2) and the generalized enstrophy representation function $\tilde{A}$ defined in (3.4) have the following properties.

(a) $\tilde{H}$ is a bounded continuous function mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$.

(b) There exists $C < \infty$ such that for each n

$$\sup_{\zeta \in \Omega_n} |\tilde{H}(Y_n(\zeta)) - H_n(\zeta)| \leqslant CK^2 \left(\frac{\log n}{n}\right)^{1/2}$$

(c) $\tilde{A}$ is a bounded continuous function mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$ and

$$\tilde{A}(Y_n(\zeta)) = A_n(\zeta) \qquad \text{for all} \quad \zeta \in \Omega_n$$

As a prelude to the proof, we recall some basic facts about the discrete Fourier transform as they pertain to our lattice model.[38] Each microstate $\zeta \in \Omega_n$ has a representation

$$\zeta(s) = \sum_{z \in \mathscr{L}^*} \hat{\zeta}(z)\, e^{2\pi i z \cdot s}$$

as a finite Fourier sum over

$$\mathscr{L}^* \doteq \{z = (z_1, z_2) \in \mathbb{Z}^2 : -2^{m-1} < z_1, z_2 \leqslant 2^{m-1}\}$$

where the coefficients $\hat{\zeta}(z)$ are given by

$$\hat{\zeta}(z) = \frac{1}{n} \sum_{s \in \mathscr{L}} \zeta(s)\, e^{-2\pi i z \cdot s}$$

Since ζ is a real-valued function on $\mathscr{L}$, it follows that $\hat{\zeta}(-z) = \hat{\zeta}(z)^*$ and $\hat{\zeta}(z + (2^m, 0)) = \hat{\zeta}(z + (0, 2^m)) = \hat{\zeta}(z)$. These definitions explain the definition of the lattice Hamiltonian H_n and the lattice Green's function g_n in Section 2. Indeed, the finite sum

$$H_n(\zeta) = \tfrac{1}{2} \sum_{0 \neq z \in \mathscr{L}^*} |2\pi z|^{-2}\, |\hat{\zeta}(z)|^2 \tag{4.4}$$

which coincides with the expression (2.8), is the spectral truncation of the Fourier series expansion for the Hamiltonian defined in (2.4); namely,

$$H(\omega) = \tfrac{1}{2} \sum_{0 \neq z \in \mathbb{Z}^2} |2\pi z|^{-2}\, |\hat{\omega}(z)|^2, \qquad \text{where} \quad \hat{\omega}(z) \doteq \int_{\mathscr{X}} \omega(x)\, e^{-2\pi i z \cdot x}\, dx$$

We now turn to the proof of Lemma 4.4. Throughout, C denotes a generic positive constant that might vary from line to line.

Proof of Lemma 4.4. (a) The boundedness and continuity of $\tilde{H}$ are deduced from the representation formula

$$\tilde{H}(\mu) = \tfrac{1}{2} \sum_{0 \neq z \in \mathbb{Z}^2} |2\pi z|^{-2} \, |\hat{\eta}(z)|^2 \tag{4.5}$$

in which

$$\hat{\eta}(z) = \hat{\eta}(\mu, z) \doteq \int_{\mathscr{X} \times \mathscr{Y}} y e^{-2\pi i z \cdot x} \mu(dx \times dy)$$

The quantities $\{\hat{\eta}(z), z \in \mathbb{Z}^2\}$ are the Fourier coefficients of the mean function

$$\eta(x) = \eta(\mu, x) \doteq \int_{\mathscr{Y}} y \, \tau(x, dy)$$

where τ is the stochastic kernel appearing in the decomposition $\mu(dx \times dy) = \theta(dx) \otimes \tau(x, dy)$. As discussed at the beginning of Subsection 3.2, any $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ admits such a decomposition with $\theta(dx) = dx$ on $\mathscr{X}$ and $\operatorname{supp} \tau(x, \cdot) \subset \mathscr{Y}$ for all $x \in \mathscr{X}$. The formula (4.5) therefore follows directly from the definition (3.2) of $\tilde{H}$.

To prove the desired bound for $\tilde{H}$, we observe that (4.5) yields

$$\tilde{H}(\mu) \leqslant C \sum_{z \in \mathbb{Z}^2} |\hat{\eta}(z)|^2 = C \int_{\mathscr{X}} \eta(x)^2 \, dx \leqslant CK^2$$

where $K \doteq \max\{|y| : y \in \mathscr{Y}\}$.

To prove the continuity of $\tilde{H}$, let $\{\mu_n, n \in [\mathrm{N}]\}$ be a sequence in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ converging weakly to μ. Then the corresponding $\hat{\eta}(\mu_n, z)$ converge to $\hat{\eta}(\mu, z)$ for every $z \in \mathbb{Z}^2$, as is clear from the definition of $\hat{\eta}$. Now the desired convergence of $\tilde{H}(\mu_n)$ to $\tilde{H}(\mu)$ follows from (4.5) by splitting this sum over z into two parts, a sum over $|z| \leqslant M$ for some suitably large M and the complementary sum. Since the complementary sum is bounded by CK^2M^{-2}, taking M as large as desired and sending $n \to \infty$ completes the proof.

(b) For $\zeta \in \Omega_n$ we introduce the piecewise constant function

$$\xi(x) = \xi(\zeta, x) \doteq \sum_{s \in \mathscr{L}} 1_{M(s)}(x) \, \zeta(s) \qquad \text{for} \quad x \in \mathscr{X}$$

and its Fourier transform

$$\hat{\xi}(z) = \int_{\mathscr{X}} \xi(x)\, e^{-2\pi i z \cdot x}\, dx = \sum_{s \in \mathscr{L}} \zeta(s) \int_{M(s)} e^{-2\pi i z \cdot x}\, dx$$

The proof that $\tilde{H}(Y_n(\zeta))$ approximates $H_n(\zeta)$ is based on the Fourier representation formula

$$\tilde{H}(Y_n(\zeta)) = \tfrac{1}{2} \sum_{0 \neq z \in \mathbb{Z}^2} |2\pi z|^{-2}\, |\hat{\xi}(z)|^2 \tag{4.6}$$

which is a consequence of (4.5) and the fact that $\hat{\eta}(Y_n(\zeta), z) = \hat{\xi}(\zeta, z)$ for $\zeta \in \Omega_n$ and $z \in \mathbb{Z}^2$. Given this expression for $\tilde{H}(Y_n(\zeta))$ and the analogous expression (4.4) for $H_n(\zeta)$, we can estimate their difference as follows:

$$\begin{aligned}
&|\tilde{H}(Y_n(\zeta)) - H_n(\zeta)| \\
&\quad \leqslant \tfrac{1}{2} \sum_{0 \neq z \in \mathscr{L}^*} |2\pi z|^{-2}\, \big||\hat{\xi}(z)|^2 - |\hat{\zeta}(z)|^2\big| + \tfrac{1}{2} \sum_{z \in \mathbb{Z}^2 \backslash \mathscr{L}^*} |2\pi z|^{-2}\, |\hat{\xi}(z)|^2 \\
&\quad \doteq E_1 + E_2
\end{aligned}$$

where we use the shorthand E_1 and E_2 to denote these two terms.

We estimate E_1 by first noting that for every $z \in \mathscr{L}^*$

$$\begin{aligned}
|\hat{\xi}(z) - \hat{\zeta}(z)| &= \left| \sum_{s \in \mathscr{L}} \zeta(s) \int_{M(s)} [e^{-2\pi i z \cdot x} - e^{-2\pi i z \cdot s}]\, dx \right| \\
&\leqslant \frac{1}{n} \sum_{s \in \mathscr{L}} K\, |2\pi z|\, [\operatorname{diam} M(s)] \\
&= CKn^{-1/2}\, |z|
\end{aligned}$$

Hence

$$\begin{aligned}
E_1 &\leqslant C \sum_{0 \neq z \in \mathscr{L}^*} |z|^{-2}\, |\hat{\xi}(z) - \hat{\zeta}(z)|\, (|\hat{\xi}(z) + \hat{\zeta}(z)|) \\
&\leqslant C \left\{ \sum_{0 \neq z \in \mathscr{L}^*} |z|^{-4}\, |\hat{\xi}(z) - \hat{\zeta}(z)|^2 \right\}^{1/2} \left\{ 2 \sum_{0 \neq z \in \mathscr{L}^*} [|\hat{\xi}(z)|^2 + |\hat{\zeta}(z)|^2] \right\}^{1/2} \\
&\leqslant CK^2 n^{-1/2} \left\{ \sum_{0 \neq z \in \mathscr{L}^*} |z|^{-2} \right\}^{1/2} \\
&\leqslant CK^2 n^{-1/2} (\log n)^{1/2}
\end{aligned}$$

which has the desired form.

We estimate E_2 in a similar fashion, obtaining

$$\begin{aligned}
E_2 &\leqslant C \sum_{z \in \mathbb{Z}^2 \setminus \mathscr{L}^*} |z|^{-2} \, |\hat{\xi}(z)|^2 \\
&\leqslant C \left\{ \sum_{z \in \mathbb{Z}^2 \setminus \mathscr{L}^*} |z|^{-4} \, |\hat{\xi}(z)|^2 \right\}^{1/2} \left\{ \sum_{z \in \mathbb{Z}^2 \setminus \mathscr{L}^*} |\hat{\xi}(z)|^2 \right\}^{1/2} \\
&\leqslant CK^2 \left\{ \sum_{z \in \mathbb{Z}^2 \setminus \mathscr{L}^*} |z|^{-4} \right\}^{1/2} \\
&\leqslant CK^2 n^{-1/2}
\end{aligned}$$

These estimates for E_1 and E_2 are both uniform over $\zeta \in \Omega_n$. Combining them gives the claimed approximation.

(c) Since a is a continuous function mapping $\mathscr{Y}$ into $\mathbb{R}$, the boundedness and continuity of $\tilde{A}$ are obvious. For $\zeta \in \Omega_n$

$$\tilde{A}(Y_n(\zeta)) = \int_{\mathscr{X} \times \mathscr{Y}} a(y) \, Y_n(\zeta, dx \times dy) = \frac{1}{n} \sum_{s \in \mathscr{L}} a(\zeta(s)) = A_n(\zeta) \quad \blacksquare$$

5. PROOF OF THE LARGE DEVIATION PRINCIPLE FOR $W_{n,r}$

In this section we outline a proof of Lemma 4.1. Specifically, we shall prove that $W_{n,r}$ satisfies the two-parameter large deviation principle on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function $R(\cdot \mid \theta \times \rho)$ with respect to the probability measures Π_n. That is, we shall establish the following:

- Large deviation upper bound. For any closed subset F of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

$$\limsup_{r \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log \Pi_n\{ W_{n,r} \in F \} \leqslant -R(F \mid \theta \times \rho)$$

- Large deviation lower bound. For any open subset G of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

$$\liminf_{r \to \infty} \liminf_{n \to \infty} \frac{1}{n} \log \Pi_n\{ W_{n,r} \in G \} \geqslant -R(G \mid \theta \times \rho)$$

This large deviation principle is equivalent to the Laplace principle stated in Lemma 4.1.

While the proofs of these large deviation bounds involve a number of technicalities, it is surprisingly easy to motivate this large deviation principle by a heuristic argument. The ease with which we can deduce the large deviation behavior of $W_{n,r}$ is the main technical advantage of our approach. We also gain a conceptual advantage from the fact that $W_{n,r}$ is a natural coarse-graining of the random vorticity field. Before proceeding with the proofs, we indicate the heuristic reasoning that suggests the large deviation behavior of $W_{n,r}$.

We write the doubly indexed process in the form

$$W_{n,r}(dx \times dy) = dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, L_{n,r,k}(dy) \tag{5.1}$$

where

$$L_{n,r,k}(dy) = L_{n,r,k}(\zeta, dy) \doteq \frac{1}{n/2^r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(dy)$$

We recall that $n = 2^{2m}$ and that r is even with $n > 2^r$. Since $D_{r,k}$ contains $n/2^r$ lattice sites s, each $L_{n,r,k}$ is an empirical measure taking values in $\mathscr{P}(\mathscr{Y})$. Since $\{\zeta(s), s \in \mathscr{L}\}$ is an independent collection of random variables with common distribution ρ, Sanov's Theorem implies that for each r and k the sequence of empirical measures $L_{n,r,k}$ satisfies the large deviation principle on $\mathscr{P}(\mathscr{Y})$ with scaling constants $n/2^r$ and rate function $R(\cdot \mid \rho)$, the relative entropy on $\mathscr{P}(\mathscr{Y})$.[9, 10, 13] In other words, for any closed subset F of $\mathscr{P}(\mathscr{Y})$

$$\limsup_{n \to \infty} \frac{1}{n/2^r} \log \Pi_n\{L_{n,r,k} \in F\} \leqslant -R(F \mid \rho)$$

and for any open subset G of $\mathscr{P}(\mathscr{Y})$

$$\liminf_{n \to \infty} \frac{1}{n/2^r} \log \Pi_n\{L_{n,r,k} \in G\} \geqslant -R(G \mid \rho)$$

Now let us suppose that $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ has finite relative entropy with respect to $\theta \times \rho$ and has the special form

$$\mu(dx \times dy) = dx \otimes \tau(x, dy), \qquad \text{where} \quad \tau(x, dy) \doteq \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \tau_k(dy) \tag{5.2}$$

and $\tau_1, \ldots, \tau_{2^r}$ are probability measures on $\mathscr{Y}$. The representation (5.1), Sanov's Theorem, and the independence of $L_{n,r,1}, \ldots, L_{n,r,2^r}$ suggest that

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log \Pi_n\{W_{n,r}\sim\mu\} &= \lim_{n\to\infty}\frac{1}{n}\log \Pi_n\{L_{n,r,k}\sim\tau_k, k=1,..., 2^r\}\\
&= \frac{1}{2^r}\sum_{k=1}^{2^r}\lim_{n\to\infty}\frac{1}{n/2^r}\log \Pi_n\{L_{n,r,k}\sim\tau_k\}\\
&\approx -\frac{1}{2^r}\sum_{k=1}^{2^r} R(\tau_k \mid \rho)\\
&= -\int_{\mathscr{X}} R(\tau(x,\cdot)\mid\rho)\, dx\\
&= -\int_{\mathscr{X}}\int_{\mathscr{Y}}\left(\log\frac{d\tau(x,\cdot)}{d\rho(\cdot)}(y)\right)\tau(x,dy)\,dx\\
&= -\int_{\mathscr{X}\times\mathscr{Y}}\left(\log\frac{d\mu}{d(\theta\times\rho)}(x,y)\right)\mu(dx\times dy)\\
&= -R(\mu\mid\theta\times\rho)
\end{aligned}$$

The two-parameter large deviation principle for $W_{n,r}$ with rate function $R(\cdot\mid\theta\times\rho)$ is therefore certainly plausible in view of the fact that any measure $\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ can be well approximated, as $r\to\infty$, by a sequence of measures of the form (5.2). This approximation property is proved in Lemma 3.2 of ref. 1.

We now prove the large deviation upper bound and lower bound for $W_{n,r}$. Our proofs will follow the above heuristic reasoning, but will replace vague statements such as $W_{n,r}\sim\mu$ by precise statements such as $W_{n,r}\in B(\mu,\varepsilon)$ and $W_{n,r}\in\bar{B}(\mu,\varepsilon)$. Here and in what follows, $B(\mu,\varepsilon)$ denotes the open ball centered at μ with radius ε and $\bar{B}(\mu,\varepsilon)$ denotes the closed ball centered at μ with radius ε. These balls are defined with respect to the dual-bounded-Lipschitz metric d in (4.2).

Proof of the Large Deviation Upper Bound for $W_{n,r}$. Since $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ is compact, a closed subset F of this space is automatically compact. By a standard covering argument, the large deviation upper bound for compact sets follows immediately from the large deviation upper bound for closed balls. We now prove the bound for closed balls.

Let $\mu=\theta\otimes\tau\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ and $\varepsilon>0$ be given. For any $n=2^{2m}$ and even r with $r<2m$, we define the closed set

$$F_{r,\varepsilon}\doteq\left\{(\nu_1,..., \nu_{2^r})\in\mathscr{P}(\mathscr{Y})^{2^r} : \theta(dx)\otimes\sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\,\nu_k(dy)\in\bar{B}(\mu,\varepsilon)\right\}$$

We also define $\mathscr{P}_{\theta,r}(\mathscr{X}\times\mathscr{Y})$ to be the set of $\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ having the form

$$\mu(dx\times dy)=\theta(dx)\otimes\nu(x,dy), \qquad \text{where} \quad \nu(x,dy)=\sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\,\nu_k(dy)$$

for some $\nu_1,\dots,\nu_{2^r}\in\mathscr{P}(\mathscr{Y})$. By Sanov's Theorem and Lemmas 2.5, 2.6, 2.7, and 2.8 in ref. 24, for each r the sequence $\{(L_{n,r,1},\dots,L_{n,r,2^r}),\ n=2^{2m}, m\in\mathbb{N}\}$ satisfies the large deviation principle on $\mathscr{P}(\mathscr{Y})^{2^r}$ with scaling constants $n/2^r$ and rate function

$$(\nu_1,\dots,\nu_{2^r})\mapsto\sum_{k=1}^{2^r}R(\nu_k\,|\,\rho)$$

Since $\theta\{D_{r,k}\}=1/2^r$,

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n/2^r}\log\Pi_n\{W_{n,r}\in\bar{B}(\mu,\varepsilon)\}\\
&\quad=\limsup_{n\to\infty}\frac{1}{n/2^r}\log\Pi_n\{(L_{n,r,1},\dots,L_{n,r,2^r})\in F_{r,\varepsilon}\}\\
&\quad\leqslant-2^r\inf\left\{\frac{1}{2^r}\sum_{k=1}^{2^r}R(\nu_k\,|\,\rho):(\nu_1,\dots,\nu_{2^r})\in F_{r,\varepsilon}\right\}\\
&\quad=-2^r\inf\left\{\int_{\mathscr{X}}R(\nu(x,\cdot)\,|\,\rho(\cdot))\,dx : dx\otimes\nu(x,dy)\in\bar{B}(\mu,\varepsilon)\cap\mathscr{P}_{\theta,r}(\mathscr{X}\times\mathscr{Y})\right\}\\
&\quad\leqslant-2^r\inf\left\{\int_{\mathscr{X}}R(\nu(x,\cdot)\,|\,\rho(\cdot))\,dx : dx\otimes\nu(x,dy)\in\bar{B}(\mu,\varepsilon)\right\}\\
&\quad=-2^rR(\bar{B}(\mu,\varepsilon)\,|\,\theta\times\rho)
\end{aligned}$$

Dividing through by 2^r and taking the limit superior as $r\to\infty$ yields the large deviation upper bound for the closed ball $\bar{B}(\mu,\varepsilon)$. ∎

Proof of the Large Deviation Lower Bound for $W_{n,r}$. To prove the large deviation lower bound, we will need the following Jensen-type inequality whose proof can be found in Lemma 3.5 in ref. 1.

Lemma 5.1. Let γ be a probability measure on $\mathscr{X}$ and τ a stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$. Then

$$\int_{\mathscr{X}}R(\tau(x,\cdot)\,|\,\rho(\cdot))\,\gamma(dx)\geqslant R\left(\int_{\mathscr{X}}\tau(x,\cdot)\,\gamma(dx)\,\Big|\,\rho(\cdot)\right)$$

Let μ be any measure in the open set G and let $\tau(x, dy)$ be the stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$ appearing in the decomposition $\mu(dx \times dy) = \theta(dx) \otimes \tau(x, dy)$. We choose $\varepsilon > 0$ so that $B(\theta \otimes \tau, \varepsilon) \subset G$. For even r and $k \in \{1,..., 2^r\}$ define the probability measures on $\mathscr{Y}$

$$\tau_k^r(\cdot) \doteq 2^r \int_{D_{r,k}} \tau(x, \cdot)\, dx$$

and the stochastic kernels on $\mathscr{Y}$ given $\mathscr{X}$

$$\tau^r(x, dy) \doteq \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \tau_k^r(dy)$$

By Lemma 3.2 in ref. 1, $\theta \otimes \tau^r \Rightarrow \theta \otimes \tau$ as $r \to \infty$. Hence there exists N such that for all even $r \geqslant N$ $B(\theta \otimes \tau^r, \varepsilon/2) \subset B(\theta \otimes \tau, \varepsilon)$. We also define the open set

$$G_{r,\varepsilon} \doteq \left\{(\nu_1,..., \nu_{2^r}) \in \mathscr{P}(\mathscr{Y})^{2^r} : \theta(dx) \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \nu_k(dy) \in B(\theta \otimes \tau^r, \varepsilon/2)\right\}$$

Then for all even $r \geqslant N$ the large deviation principle satisfied by $\{(L_{n,r,1},..., L_{n,r,2^r}),\ n = 2^{2m},\ m \in \mathbb{N}\}$ and Lemma 5.1 imply that

$$\begin{aligned}
&\liminf_{n \to \infty} \frac{1}{n} \log \Pi_n\{W_{n,r} \in G\} \\
&\quad \geqslant \liminf_{n \to \infty} \frac{1}{n} \log \Pi_n\{W_{n,r} \in B(\theta \otimes \tau^r, \varepsilon/2)\} \\
&\quad = \frac{1}{2^r} \liminf_{n \to \infty} \frac{1}{n/2^r} \log \Pi_n\{(L_{n,r,1},..., L_{n,r,2^r}) \in G_{r,\varepsilon}\} \\
&\quad \geqslant -\frac{1}{2^r} \inf\left\{\sum_{k=1}^{2^r} R(\nu_k \mid \rho) : (\nu_1,..., \nu_{2^r}) \in G_{r,\varepsilon}\right\} \geqslant -\frac{1}{2^r} \sum_{k=1}^{2^r} R(\tau_k^r \mid \rho) \\
&\quad = -\frac{1}{2^r} \sum_{k=1}^{2^r} R\left(2^r \int_{D_{r,k}} \tau(x, \cdot)\, dx \,\middle|\, \rho(\cdot)\right) \\
&\quad \geqslant -\sum_{k=1}^{2^r} \int_{D_{r,k}} R(\tau(x, \cdot) \mid \rho(\cdot))\, dx \\
&\quad = -\int_{\mathscr{X}} R(\tau(x, \cdot) \mid \rho(\cdot))\, dx = -R(\mu \mid \theta \times \rho)
\end{aligned}$$

The last equality is a consequence of the chain rule (ref. 13, Cor. C.3.2). It follows that

$$\liminf_{r \to \infty} \liminf_{n \to \infty} \frac{1}{n} \log \Pi_n\{W_{n,r} \in G\} \geqslant -R(\mu \mid \theta \times \rho)$$

Since $\mu \in G$ is arbitrary, the proof of the large deviation lower bound for the open set G is complete. ▮

ACKNOWLEDGMENTS

Research of R.S.E. was supported in part by grants from the Department of Energy (DE-FG02-99ER25376) and the National Science Foundation (NSF-DMS-9700852). The research of B.T. was supported in part by grants from the Department of Energy (DE-FG02-99ER25376) and the National Science Foundation (NSF-DMS-9971204).

REFERENCES

1. 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).
2. H. Brands, J. Stulemeyer, and R. A. Pasmanter, A mean field prediction of the asymptotic state of decaying 2D turbulence, *Phys. Fluids A* **9**:2815 (1997).
3. F. P. Bretherton and D. B. Haidvogel, Two-dimensional turbulence over topography, *J. Fluid Mech.* **78**:129–154 (1976).
4. E. Caglioti, P. L. Lions, C. Marchioro, and M. Pulvirenti, A special class of stationary flows for two-dimensional Euler equations: A statistical mechanics description, *Commun. Math. Phys.* **143**:501–525 (1992).
5. P. H. Chavanis and J. Sommeria, Classification of self-organized structures in two-dimensional turbulence: The case of a bounded domain, *J. Fluid Mech.* **314**:267–297 (1996).
6. A. J. Chorin, *Vorticity and Turbulence* (Springer-Verlag, New York, 1994).
7. A. J. Chorin, Partition functions and equilibrium measures in two-dimensional and quasi-three-dimensional turbulence, *Phys. Fluids* **8**:2656–2660 (1996).
8. A. J. Chorin, A. Kast, and R. Kupferman, Optimal prediction of underresolved dynamics, *Proc. Nat. Acad. Sci. USA* **95**:4094–4098 (1998).
9. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Their Applications* (Jones & Bartlett, Boston, 1993).
10. J.-D. Deuschel and D. W. Stroock, *Large Deviations* (Academic Press, Boston, 1989).
11. M. DiBattista, A. Majda, and B. Turkington, Prototype geophysical vortex structures via large-scale statistical theory, *Geophys. Astrophys. Fluid Dyn.* **89**:235–283 (1998).
12. R. M. Dudley, *Real Analysis and Probability* (Wadsworth & Brooks/Cole, Pacific Grove, California, 1989).
13. P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations* (Wiley, New York, 1997).

14. T. Eisele and R. S. Ellis, Symmetry breaking and random waves for magnetic systems on a circle, *Z. Wahrsch. verw. Geb.* **63**:279–348 (1983).
15. R. S. Ellis, An overview of the theory of large deviations and applications to statistical mechanics, *Scand. Actuarial J.* **1**:97–142 (1995).
16. R. S. Ellis, K. Haven, and B. Turkington, Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles, University of Massachusetts, technical report, 1999.
17. 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).
18. D. Holm, J. Marsden, T. Ratiu, and A. Weinstein, Nonlinear stability of fluid and plasma equilibria, *Rev. Mod. Phys.* **123**:1–116 (1985).
19. A. D. Ioffe and V. M. Tihomirov, *Theory of Extremal Problems* (Elsevier, North-Holland, New York, 1979).
20. B. Jeuttner and A. Thess, On the symmetry of self-organized structures in two-dimensional turbulence, *Phys. Fluids A* **7**:2108–2110 (1995).
21. G. Joyce and D. C. Montgomery, Negative temperature states for the two-dimensional guiding center plasma, *J. Plasma Phys.* **10**:107–121 (1973).
22. M. Kiessling, Statistical mechanics of classical particles with logarithmic interactions, *Commun. Pure Appl. Math.* **46**:27–56 (1993).
23. R. Kraichnan, Statistical dynamics of two-dimensional flow, *J. Fluid Mech.* **67**:155–175 (1975).
24. J. Lynch and J. Sethuraman, Large deviations for processes with independent increments, *Ann. Probab.* **15**:610–627 (1987).
25. C. Marchioro and M. Pulvirenti, *Mathematical Theory of Incompressible Nonviscous Fluids*, Appl. Math. Sci., Vol. 96 (Springer-Verlag, New York, 1994).
26. J. C. McWilliams, The emergence of isolated coherent vortices in turbulent flow, *J. Fluid Mech.* **146**:21–43 (1984).
27. 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).
28. J. Miller, Statistical mechanics of Euler equations in two dimensions, *Phys. Rev. Lett.* **65**:2137–2140 (1990).
29. 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).
30. D. Montgomery and G. Joyce, Statistical mechanics of negative temperature states, *Phys. Fluids* **17**:1139–1145 (1974).
31. D. Montgomery, W. T. Matthaeus, D. Martinez, and S. Oughton, Relaxation in two dimensions and the Sinh–Poisson equation, *Phys. Fluids A* **4**:3–6 (1992).
32. L. Onsager, Statistical hydrodynamics, *Suppl. Nuovo Cim.* **6**:279–287 (1949).
33. R. Robert, Concentration et entropie pour les mesures d'Young, *C. R. Acad. Sci. Paris Ser. I* **309**:757–760 (1989).
34. R. Robert, A maximum-entropy principle for two-dimensional perfect fluid dynamics, *J. Statist. Phys.* **65**:531–553 (1991).
35. R. Robert and J. Sommeria, Statistical equilibrium states for two-dimensional flows, *J. Fluid Mech.* **229**:291–310 (1991).
36. P. Santangelo, R. Benzi, and B. Legras, The generation of vortices in high-resolution, two-dimensional decaying turbulence and the influence of initial conditions on the breaking of self-similarity, *Phys. Fluids A* **1**:1027–1034 (1989).
37. E. Segre and S. Kida, Late states of incompressible 2D decaying vorticity fields, *Fluid Dyn. Res.* **23**:89–112 (1998).
38. J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis* (Springer-Verlag, New York, 1980).

39. B. Turkington, Statistical equilibrium measures and coherent states in two-dimensional turbulence, *Commun. Pure Appl. Math*. **52**:781–809 (1999).
40. B. Turkington and N. Whitaker, Statistical equilibrium computations of coherent structures in turbulent shear layers, *SIAM J. Sci. Comput*. **17**:1414–1433 (1996).
41. N. Whitaker and B. Turkington, Maximum entropy states for rotating vortex patches, *Phys. Fluids A* **6**:3963–3973 (1994).
42. E. Zeidler, *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization* (Springer-Verlag, New York, 1985).