The Annals of Probability
1999, Vol. 27, No. 1, 297–324

SPATIALIZING RANDOM MEASURES: DOUBLY INDEXED PROCESSES AND THE LARGE DEVIATION PRINCIPLE

BY CHRISTOPHER BOUCHER, RICHARD S. ELLIS[1] AND BRUCE TURKINGTON[2]

Illinois Wesleyan University, University of Massachusetts and University of Massachusetts

The main theorem is the large deviation principle for the doubly indexed sequence of random measures

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

Here θ is a probability measure on a Polish space $\mathscr{X}$, $\{D_{r,k}, k = 1, \ldots, 2^r\}$ is a dyadic partition of $\mathscr{X}$ (hence the use of 2^r summands) satisfying $\theta\{D_{r,k}\} = 1/2^r$ and $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ is an independent, identically distributed sequence of random probability measures on a Polish space $\mathscr{Y}$ such that $\{L_{q,k}, q \in \mathbb{N}\}$ satisfies the large deviation principle with a convex rate function. A number of related asymptotic results are also derived.

The random measures $W_{r,q}$ have important applications to the statistical mechanics of turbulence. In a companion paper, the large deviation principle presented here is used to give a rigorous derivation of maximum entropy principles arising in the well-known Miller–Robert theory of two-dimensional turbulence as well as in a modification of that theory recently proposed by Turkington.

1. Introduction. Statistical mechanics and probability theory thrive in a mutually beneficial, symbiotic relationship. Since the time of the founders, Boltzmann and Gibbs, statistical mechanics has stimulated research into random phenomena which has enriched probability theory immeasurably. For example, ergodic theory, the theory of large deviations and the theory of interacting particle systems all owe their origins to statistical mechanics, which continues to be an important source of problems in these areas. Conversely, advances in probability theory—in particular, in the theory of large deviations—continue to yield insights into ever more complicated statistical mechanical models.

One of the main contributions of the theory of large deviations to statistical mechanics is the systematization of a procedure for the asymptotic evaluation of key statistical mechanical quantities in terms of variational formulas over sets of macrostates. These applications, well known in the study of spin systems such as the Ising model, are explained in that context in [8] as well as in numerous other references. Recent applications arising in the statistical

Received April 1998.
[1]Supported in part by NSF Grant DMS-97-00852.
[2]Supported in part by NSF Grant DMS-96-00060.
AMS 1991 *subject classifications*. Primary 60F10; secondary 82B99.
Key words and phrases. Large deviation principle, doubly indexed processes, random measures, Sanov's theorem, turbulence.

mechanics of turbulence can also be treated via the theory of large deviations in a related but novel way. The analysis of such statistical mechanical models is greatly facilitated by using a class of doubly indexed processes, whose large deviation principles it is the aim of this paper to derive. The double indexing is not merely a mathematical contrivance, but reflects fundamental multiscale aspects of the models under consideration.

In order to put flesh on this skeleton of assertions, let us consider a special case of the doubly indexed processes with which we will deal. This process arises in the analysis of a specific model of two-dimensional turbulence that will be studied in [2]. The asymptotics of this model lead to a rigorous derivation of maximum entropy principles arising in the well-known Miller–Robert theory of two-dimensional turbulence [16, 18, 19] as well as in a modification of that theory recently proposed by Turkington [22]. Let T^2 denote the unit square $[0,1)\times[0,1)$ with periodic boundary conditions. For u and v in $\mathbb{N}$ we set $r \doteq 2u$, $q \doteq 2^{2v}$ and $n \doteq 2^r q$ and consider a regular dyadic partition of T^2 into 2^r squares $D_{r,k}$, called "macrocells," each having area $1/2^r$. We also consider a regular dyadic partition of each $D_{r,k}$ into $q = n/2^r$ squares, called "microcells," each having area $1/n$. The model is defined on the sites of the uniform lattice $\mathscr{L}$ of n points in T^2 containing the origin and having intersite spacing equal to $1/n^{1/2}$ in each coordinate direction. Each macrocell $D_{r,k}$ contains $q = n/2^r$ points of $\mathscr{L}$ and each microcell one point of $\mathscr{L}$. For $s \in \mathscr{L}$, $M(s)$ denotes the unique microcell containing s. Let ρ be a probability measure on $\mathbb{R}$ with bounded support $\mathscr{Y}$. The configuration space of the model is the product space $\Omega_n \doteq \mathscr{Y}^n$, and a typical configuration is denoted by $\zeta = \{\zeta(s), s \in \mathscr{L}\}$, which is referred to as the vorticity field. We denote by P_n the finite product measure on Ω_n which assigns to a Borel subset B of Ω_n the probability

$$P_n\{B\} \doteq \int_B \prod_{s\in\mathscr{L}} \rho(d\zeta(s)).$$

With respect to P_n, the coordinates $\zeta(s)$ are i.i.d. random variables with common distribution ρ.

Write $\theta(dx)$ or dx for Lebesgue measure on T^2 and define $\mathscr{P}_\theta(T^2\times\mathscr{Y})$ to be the set of probability measures on $T^2\times\mathscr{Y}$ having first marginal θ. We consider the random probability measure

$$W_{r,q}(dx\times dy) = W_{r,q}(\zeta, dx\times dy) \doteq dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, L_{q,k}(dy), \tag{1.1}$$

where $L_{q,k}$ is the empirical measure

$$L_{q,k}(dy) = L_{q,k}(\zeta, dy) \doteq \frac{1}{q} \sum_{s\in\mathscr{L}\cap D_{r,k}} \delta_{\zeta(s)}(dy).$$

Thus $W_{r,q}$ assigns to a Borel subset B of $T^2\times\mathscr{Y}$ the probability

$$W_{r,q}\{B\} \doteq \sum_{k=1}^{2^r} \int_B 1_{D_{r,k}}(x)\, dx\, L_{q,k}(dy).$$

$W_{r,q}$ takes values in $\mathscr{P}_\theta(T^2 \times \mathscr{Y})$. The large deviation principle satisfied by $W_{r,q}$ as well as by a large class of generalizations is stated in Theorem 2.4. These generalizations are obtained by replacing T^2 and $\mathscr{Y}$ by arbitrary Polish spaces and $L_{q,k}$ by other random measures on $\mathscr{Y}$. Theorem 2.4 is proved in Section 3.

The following heuristic calculation motivates the large deviation principle for $W_{r,q}$ and suggests the form of the rate function in this particular case. The proofs of the large deviation lower bound and upper bound in the general context of Theorem 2.4 will follow this line of reasoning. We denote by $R(\cdot\,|\,\cdot)$ the relative entropy and by $\mathscr{P}(\mathscr{Y})$ the set of probability measures on $\mathscr{Y}$. Let $\tau_1, \dots, \tau_{2^r}$ be probability measures on $\mathscr{Y}$ and suppose that $\mu \in \mathscr{P}_\theta(T^2 \times \mathscr{Y})$ satisfies $R(\mu \,|\, \theta \times \rho) < \infty$ and has the form

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

By Sanov's theorem, for each k $\{L_{q,k}, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with rate function $R(\cdot\,|\,\rho)$. Since $L_{q,1}, L_{q,2}, \dots, L_{q,2^r}$ are independent,

$$\begin{aligned}
&\lim_{q\to\infty} \frac{1}{2^r q} \log P_n\{W_{r,q} \sim \mu\} \\
&\quad = \lim_{q\to\infty} \frac{1}{2^r q} \log P_n\{L_{q,1} \sim \tau_1, L_{q,2} \sim \tau_2, \dots L_{q,2^r} \sim \tau_{2^r}\} \\
&\quad = \frac{1}{2^r} \sum_{k=1}^{2^r} \lim_{q\to\infty} \frac{1}{q} \log P_n\{L_{q,k} \sim \tau_k\} \\
&\quad \approx -\frac{1}{2^r} \sum_{k=1}^{2^r} R(\tau_k \,|\, \rho) = -\sum_{k=1}^{2^r} \int_{D_{r,k}} R(\tau(x,\cdot) \,|\, \rho(\cdot))\, dx \qquad (1.2) \\
&\quad = -\int_{T^2} R\big(\tau(x,\cdot) \,|\, \rho(\cdot)\big)\, dx \\
&\quad = -\int_{T^2} \int_{\mathscr{Y}} \left(\log \frac{d\tau(x,\cdot)}{d\rho(\cdot)}(y)\right) \tau(x, dy)\, dx \\
&\quad = -\int_{T^2} \int_{\mathscr{Y}} \left(\log \frac{d\mu}{d(\theta \times \rho)}(x, y)\right) \mu(dx \times dy) = -R(\mu | \theta \times \rho).
\end{aligned}$$

By Lemma 3.2, any measure $\mu \in \mathscr{P}_\theta(T^2 \times \mathscr{Y})$ can be well approximated by a sequence of measures of the form $dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \tau_k(dy)$ as $r \to \infty$. Hence, the calculation in the last display makes it reasonable to expect that $W_{r,q}$ satisfies a "two-parameter LDP," which we summarize by the notation

$$\lim_{r\to\infty} \lim_{q\to\infty} \frac{1}{2^r q} \log P_n\big\{W_{r,q} \in \cdot\big\} = -R(\cdot \,|\, \theta \times \rho). \qquad (1.3)$$

Such doubly indexed processes $W_{r,q}$, of considerable interest in their own right, are doubly interesting because of their applications to the statistical mechanics of turbulence. In order to explain this, it is useful to outline in some detail a systematic procedure, alluded to in the second paragraph of this section, for applying the theory of large deviations to the asymptotic evaluation of key statistical mechanical quantities. Whether explicitly stated or not, this procedure is at the heart of numerous analyses of statistical mechanical models. The procedure applies to spin systems such as the Ising model, to models of turbulence such as will be considered in [2] and to numerous other models.

We consider a statistical mechanical model that is defined in terms of the following data.

1. A sequence of configuration spaces $\{\Omega_n, n \in \mathbb{N}\}$.
2. A Hamiltonian $H_n(\zeta)$ of $\zeta \in \Omega_n$ and an additional function $A_n(\zeta)$ of $\zeta \in \Omega_n$. In the case of spin systems A_n could represent the interactions of the spins with an external magnetic field, while in the case of turbulence it could represent a generalized enstrophy [2]. $A_n = 0$ is allowed.
3. A sequence of positive scaling constants $b_n \to \infty$.
4. A probability measure P_n on Ω_n.

In terms of these quantities we define for each $n \in \mathbb{N}$ and $\beta \in \mathbb{R}$, the partition function

$$Z_n(\beta) \doteq \int_{\Omega_n} \exp\left[-\beta H_n(\zeta) - A_n(\zeta)\right] P_n(d\zeta)$$

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

$$P_{n,\beta}\{B\} \doteq \frac{1}{Z_n(\beta)} \int_B \exp\left[-\beta H_n(\zeta) - A_n(\zeta)\right] P_n(d\zeta).$$

For $\beta \in \mathbb{R}$ we also consider the limit

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

if it exists. The function $-\beta^{-1}\varphi(\beta)$ is known as the specific Gibbs free energy for the model.

In order to carry out a large deviation analysis of the model, the following four items are needed.

1. A Polish space $\mathscr{S}$, called the *hidden space*.
2. For each $n \in \mathbb{N}$ a random variable Y_n mapping Ω_n into $\mathscr{S}$. The sequence $\{Y_n, n \in \mathbb{N}\}$ is called the *hidden process*.
3. Bounded continuous functions $\tilde{H}$ and $\tilde{A}$ mapping $\mathscr{S}$ into $\mathbb{R}$ such that

$$H_n(\zeta) = b_n \tilde{H}(Y_n(\zeta)) + o(b_n) \quad \text{and} \quad A_n(\zeta) = b_n \tilde{A}(Y_n(\zeta)) + o(b_n) \quad \text{uniformly for } \zeta \in \Omega_n. \tag{1.4}$$

$\tilde{H}$ and $\tilde{A}$ are called *representation functions*.

4. A rate function J on $\mathscr{S}$ such that the sequence of P_n-distributions of Y_n satisfies the *large deviation principle* on $\mathscr{S}$ with scaling constants b_n and rate function J. In other words, J maps $\mathscr{S}$ into $[0, \infty]$, J has compact level sets and for any closed subset F of $\mathscr{S}$ and open subset G of $\mathscr{S}$

$$\limsup_{n\to\infty} \frac{1}{b_n} \log P_n\{Y_n \in F\} \le -J(F),$$
$$\liminf_{n\to\infty} \frac{1}{b_n} \log P_n\{Y_n \in G\} \ge -J(G),$$

where $J(B)$ denotes the infimum of J over a set B.

Given these items, the asymptotic behavior of the model is readily determined. To see this, let us summarize the large deviation principle for the P_n-distributions of Y_n by the formal notation

$$P_n(Y_n \in dz) \asymp \exp[-b_n J(z)]\, dz.$$

Substituting this into the definition of Z_n and using (1.4), one is led, by analogy with Laplace's method for integrals on $\mathbb{R}$, to the following formal limit, which is not difficult to justify:

$$\begin{aligned}
\varphi(\beta) &\doteq \lim_{n\to\infty} \frac{1}{b_n} \log Z_n(\beta) \\
&= \lim_{n\to\infty} \frac{1}{b_n} \log \int_{\Omega_n} \exp\bigl[-b_n\bigl(\beta \tilde{H}(Y_n) + \tilde{A}(Y_n)\bigr)\bigr]\, dP_n \\
&= \lim_{n\to\infty} \frac{1}{b_n} \log \int_{\mathscr{S}} \exp\bigl[-b_n\bigl(\beta \tilde{H}(z) + \tilde{A}(z)\bigr)\bigr] P_n(Y_n \in dz) \\
&\text{“=”} \lim_{n\to\infty} \frac{1}{b_n} \log \int_{\mathscr{S}} \exp\bigl[-b_n\bigl(\beta \tilde{H}(z) + \tilde{A}(z) + J(z)\bigr)\bigr]\, dz \\
&\text{“=”} \sup_{z\in\mathscr{S}} \bigl\{-\beta \tilde{H}(z) - \tilde{A}(z) - J(z)\bigr\}.
\end{aligned}$$

Similar considerations motivate the fact that for each β the sequence of $P_{n,\beta}$-distributions of the hidden process Y_n satisfies the large deviation principle on $\mathscr{S}$ with scaling constants b_n and rate function

$$J_\beta(z) \doteq J(z) + \beta \tilde{H}(z) + \tilde{A}(z) - \inf_{y\in\mathscr{S}} \bigl\{J(y) + \beta \tilde{H}(y) + \tilde{A}(y)\bigr\}.$$

It follows that if B is a Borel subset of $\mathscr{S}$ whose closure $\bar{B}$ has empty intersection with the 0-level set $\mathscr{E}_\beta \doteq \{z \in \mathscr{S}: J_\beta(z) = 0\}$, then $J_\beta(\bar{B}) > 0$, and so by the large deviation upper bound

$$P_{n,\beta}\{Y_n \in B\} \le C \exp\bigl(-b_n J_\beta(\bar{B})/2\bigr) \to 0.$$

This in turn leads to the identification of $\mathscr{S}$ as the set of possible macrostates for the model and of $\mathscr{E}_\beta$ as the set of equilibrium macrostates.

In the case of many important spin systems, the hidden space, the hidden process, the representation functions, and the large deviation principle are well known. For example, as explained in [10], in the case of the Curie–Weiss model, $\mathscr{S}$ equals $\mathbb{R}$, Y_n equals the sample mean of the spins, and the large deviation principle is given by Cramér's theorem; in the case of the Curie–Weiss–Potts model, $\mathscr{S}$ equals the set of probability vectors in $\mathbb{R}^Q$ for some $Q \in \mathbb{N}$, Y_n equals the empirical vector of the spins, and the large deviation principle is given by Sanov's theorem; in the case of the D-dimensional Ising model, $\mathscr{S}$ equals the set of strictly stationary probability measures on $\{1, -1\}^{\mathbb{Z}^D}$, Y_n equals the empirical field of the spins, and the large deviation principle is proved in [13, 17]. In any model such as these for which the hidden space, the hidden process, and the representation functions can be identified and the large deviation principle proved, the asymptotic behavior of the model can be determined as discussed in the preceding paragraph.

Let us now turn to the model of two-dimensional turbulence that will be studied in [2] and that is defined on the uniform lattice $\mathscr{L}$ of T^2 as described in the second paragraph of this section. For a standard choice of Hamiltonian H_n and generalized enstrophy A_n given in [2], the "simplest" choice of hidden process is the sequence of random measures,

$$Y_n(dx \times dy) = Y_n(\zeta, dx \times dy) \doteq dx \otimes \sum_{s \in \mathscr{L}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy). \tag{1.5}$$

In this case the hidden space is $\mathscr{P}_\theta(T^2 \times \mathscr{Y})$, and the representation functions are readily determined. On the other hand, the large deviation principle is by no means obvious. Presenting a proof that relies on complicated convex analysis, some of the details of which are not spelled out, the paper [15] states the large deviation principle for the sequence of P_n-distributions of Y_n in (1.5) with rate function $R(\cdot|\theta \times \rho)$. For this problem, the use of convex analysis seems to be the wrong tool since it does not take advantage of the relatively simple spatial dependence of Y_n and gives little insight into why the large deviation principle should hold. The proof in [15] reduces to the standard convex analysis proof of Sanov's theorem ([3], Section 6.2) if T^2 is replaced by a single point.

This state of affairs, coupled with the importance of the application to two-dimensional turbulence, led us to develop the two-parameter techniques elucidated in the present paper and applied in [2]. Our approach is to prove the requisite large deviation principle for the hidden process Y_n, not by convex analysis, but by approximating Y_n by the random measures $W_{r,q}$ in (1.1) and applying the almost intuitive two-parameter large deviation principle for $W_{r,q}$ summarized in (1.3). The approximation is straightforward. We first rewrite Y_n in the form

$$Y_n(dx \times dy) = dx \otimes \sum_{k=1}^{2^r} \sum_{s \in \mathscr{L} \cap D_{r,k}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy).$$

Replacing, for each $s \in \mathscr{L} \cap D_{r,k}$, the point mass $\delta_{\zeta(s)}$ by the average $q^{-1} \cdot \sum_{s\in\mathscr{L}\cap D_{r,k}} \delta_{\zeta(s)} = L_{q,k}$, one expects that for large q and large r

$$\begin{aligned} Y_n(dx \times dy) &\approx dx \otimes \sum_{k=1}^{2^r} \biggl(\sum_{s\in\mathscr{L}\cap D_{r,k}} 1_{M(s)}(x) \biggr) L_{q,k}(dy) \\ &= dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, L_{q,k}(dy) = W_{r,q}(dx \times dy). \end{aligned}$$

In fact, by taking r large enough, one can show that with respect to an appropriate metric on $\mathscr{P}_\theta(T^2 \times \mathscr{Y})$ the distance between Y_n and $W_{r,q}$ can be made as small as desired uniformly in q.

The large deviation principle for a wide class of generalizations of $W_{r,q}$ is formulated in Section 2 of this paper together with related asymptotic results. The main theorem is proved in Section 3. Section 4 is devoted to a restatement of the large deviation principle that is needed in [2] and other applications.

2. Statement of the large deviation theorem and examples. In this section we formulate the large deviation principle (LDP) and related asymptotic results for extensive generalizations of the random measures $W_{r,q}$ defined in (1.1). Let $(\Omega, \mathscr{F}, P)$ be a probability space, $\mathscr{Y}$ a Polish space, $\mathscr{P}(\mathscr{Y})$ the space of probability measures on $\mathscr{Y}$ with the topology of weak convergence and I a convex rate function on $\mathscr{P}(\mathscr{Y})$. The convexity of I is a natural hypothesis satisfied in many cases of interest. We assume that $\{L_q, q \in \mathbb{N}\}$ is a sequence of random variables mapping Ω into $\mathscr{P}(\mathscr{Y})$ which satisfies the large deviation principle with rate function I. Thus I maps $\mathscr{P}(\mathscr{Y})$ into $[0,\infty]$; for each $M \in [0,\infty)$, $\{\gamma \in \mathscr{P}(\mathscr{Y})\colon I(\gamma) \le M\}$ is compact (compact level sets); for any closed subset F of $\mathscr{P}(\mathscr{Y})$

$$\limsup_{q\to\infty} \frac{1}{q} \log P\{L_q \in F\} \le -I(F)$$

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

$$\liminf_{q\to\infty} \frac{1}{q} \log P\{L_q \in G\} \ge -I(G),$$

where $I(B)$ denotes the infimum of I over the set B. A basic example of such a sequence, and the one that appears in the application to two-dimensional turbulence in the companion paper [2], is the sequence L_q of empirical measures of i.i.d. random variables ζ_i taking values in $\mathscr{Y}$; thus, $L_q \doteq q^{-1} \sum_{i=1}^q \delta_{\zeta_i}$. This and other examples will be discussed in Example 2.7. We also introduce, for each $r \in \mathbb{N}$, a sequence $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ of 2^r independent random variables mapping Ω into $\mathscr{P}(\mathscr{Y})$, each having the same distribution as L_q. Finally, let $\mathscr{X}$ be a Polish space, θ a probability measure on $\mathscr{X}$ and $\Xi_r \doteq \{D_{r,k}, k = 1, \ldots, 2^r\}$ a dyadic partition of $\mathscr{X}$ satisfying the following condition.

CONDITION 2.1. For each $r \in \mathbb{N}$:

(i) $\theta\{D_{r,k}\} = 1/2^r$;
(ii) Ξ_{r+1} is a refinement of Ξ_r in the sense that $D_{r,k} = D_{r+1,2k-1} \cup D_{r+1,2k}$;
(iii) $\lim_{r\to\infty} \max_{k\in\{1,\dots,2^r\}} \operatorname{diam}(D_{r,k}) = 0$;
(iv) $\theta\{\partial D_{r,k}\} = 0$.

Part (i) of this condition states that Ξ_r is an equivolume partition. Part (ii) is needed in order to prove Lemma 3.2, which uses a martingale argument to derive a key approximation property of certain measures. In applications to turbulence, part (ii) reflects the natural way of constructing a sequence of lattice models, where each lattice is a refinement of its predecessor. We use part (iii) of Condition 2.1 to prove Lemma 3.1, which allows us to approximate an arbitrary closed set in $\mathscr{X}$ by sets in the σ-fields generated by the partitions Ξ_r. Finally, part (iv) of Condition 2.1 is needed in Section 3.5 to prove that the function J defined in Definition 2.3 is lower semicontinuous.

Let θ be Lebesgue measure on $\mathscr{X} \doteq [0, 1)$ with periodic boundary conditions. For $r \in \mathbb{N}$, taking $D_{r,k} \doteq [(k-1)/2^r, k/2^r)$, $1 \le k \le 2^r$, gives an example of a partition satisfying Condition 2.1.

REMARK 2.2. By Lemma 3.3, the assumptions that $L_{q,1}, \dots, L_{q,2^r}$ are i.i.d. copies of L_q and that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with rate function I guarantee that for each r the sequence $\{(L_{q,1}, \dots, L_{q,2^r}), q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})^{2r}$ with rate function

$$(\nu_1, \dots, \nu_{2^r}) \mapsto \sum_{k=1}^{2^r} I(\nu_k).$$

All that is needed in the sequel is that for each r $\{(L_{q,1}, \dots, L_{q,2^r}), q \in \mathbb{N}\}$ satisfies this LDP and that for each q, $L_{q,1}, \dots, L_{q,2^r}$ have the same distributions as L_q but need not be independent. While this LDP is true under much weaker hypotheses on $L_{q,k}$, we have adopted these assumptions to avoid overcomplicating the exposition.

The process whose asymptotics we wish to analyze is the doubly indexed sequence of random probability measures on $\mathscr{X} \times \mathscr{Y}$ given by

$$W_{r,q}(dx \times dy) \doteq \theta(dx) \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, L_{q,k}(dy). \tag{2.1}$$

$W_{r,q}$ maps Ω into $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ and assigns to a Borel subset B of $\mathscr{X} \times \mathscr{Y}$ the probability

$$W_{r,q}\{B\} \doteq \sum_{k=1}^{2^r} \int_B 1_{D_{r,k}}(x)\theta(dx) L_{q,k}(dy).$$

The sum in (2.1) defines a stochastic kernel $\tau(x, dy)$ on $\mathscr{Y}$ given $\mathscr{X}$. In other words, $\tau(x, dy)$ is a family of probability measures on $\mathscr{Y}$ indexed by $x \in \mathscr{X}$

and for each Borel subset B of $\mathscr{Y}$ the mapping

$$x \mapsto \tau(x, B) = \sum_{k=1}^{2^r} 1_{D_{r,k}}(x) L_{q,k}\{B\}$$

is measurable. In order to avoid the complications involved in working with a space of stochastic kernels, the measure θ has been included in the definition of $W_{r,q}$. We denote by $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ the closed subset of $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ consisting of measures with first marginal equal to θ. Then $W_{r,q}$ takes values in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. In Section 4 we introduce a standard metric that makes $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ and $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ Polish spaces.

The LDP satisfied by $W_{r,q}$ is stated in Theorem 2.4. The formal calculation given in (1.2) can easily be generalized to motivate this LDP and suggest the form of the rate function, which we next define. For any $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ there exists a stochastic kernel $\tau(x, dy)$ on $\mathscr{Y}$ given $\mathscr{X}$ such that $\mu(dx \times dy) = \theta(dx) \otimes \tau(x, dy)$ ([6], Theorem A.5.4). The definition of the rate function for $W_{r,q}$ uses this decomposition, which we summarize as $\mu = \theta \otimes \tau$.

DEFINITION 2.3. *Let I denote the convex rate function for $\{L_q\}$ on $\mathscr{P}(\mathscr{Y})$. Given $\mu = \theta \otimes \tau \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, define*

$$J(\mu) \doteq \int_{\mathscr{X}} I(\tau(x, \cdot))\, \theta(dx).$$

$J(\mu)$ is well defined since the mapping $x \in \mathscr{X} \mapsto \tau(x, \cdot) \in \mathscr{P}(\mathscr{Y})$ is measurable ([6], Theorem A.5.2) and I is nonnegative and lower semicontinuous. Clearly J is nonnegative, and because I is convex, J is convex. We will prove in general, by an indirect argument, that J has compact level sets. It is possible to see this directly in several cases. For example, assume that $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ are i.i.d. copies of the empirical measure $L_q \doteq q^{-1} \sum_{i=1}^{q} \delta_{\zeta_i}$ of i.i.d. random variables ζ_i having the common distribution ρ. Then

$$J(\mu) = R(\mu \mid \theta \times \rho),$$

where $R(\cdot|\cdot)$ denotes the relative entropy. Thus J has compact level sets since the relative entropy has this property. Details are given in Example 2.7(a).

We now state the two-parameter large deviation theorem for $W_{r,q}$. It is proved in Section 3. The convexity of the rate function I is needed in the proof (see Lemmas 3.4 and 3.5). We have no example of a nonconvex I for which the two-parameter large deviation principle in Theorem 2.4 is false.

THEOREM 2.4. *Let $W_{r,q}$ be defined by* (2.1). *We assume that $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ are i.i.d. copies of L_q, that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with a convex rate function I and that the partitions $\Xi_r = \{D_{r,k}\}$ satisfy Condition* 2.1. *Then the function J defined in Definition* 2.3 *is a convex rate function. Furthermore, the sequence $W_{r,q}$ satisfies the two-parameter LDP on*

$\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ with rate function J in the following sense. For any closed subset F of $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$

$$\limsup_{r\to\infty}\limsup_{q\to\infty}\frac{1}{2^r q}\log P\{W_{r,q}\in F\}\le -J(F),$$

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

$$\liminf_{r\to\infty}\liminf_{q\to\infty}\frac{1}{2^r q}\log P\{W_{r,q}\in G\}\ge -J(G).$$

The proof of the equivalence between the one-parameter LDP and the Laplace principle, given in Section 1.2 of [6], carries over with obvious modifications to the two-parameter setting. The two-parameter Laplace principle implied by the LDP in Theorem 2.4 is stated next. It will be needed in the application to two-dimensional turbulence given in [2].

COROLLARY 2.5. *Under the same conditions as in Theorem 2.4, the sequence $W_{r,q}$ satisfies the two-parameter Laplace principle on $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ with rate function J in the following sense. For any bounded continuous function h mapping $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ into $\mathbb{R}$*

$$\lim_{r\to\infty}\lim_{q\to\infty}\frac{1}{2^r q}\log\int_{\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\exp\big[2^r q\, h(\mu)\big]P(W_{r,q}\in d\mu) = \sup_{\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\big\{h(\mu)-J(\mu)\big\}.$$

Another corollary of Theorem 2.4 that will also be needed in the application to two-dimensional turbulence is the LDP for a sequence of measures on $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ defined in terms of the distributions of $W_{r,q}$. Let Φ be a bounded continuous function mapping $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ into $\mathbb{R}$ and consider the sequence of probability measures $\{\tilde{P}_{r,q}, r\in\mathbb{N}, q\in\mathbb{N}\}$ on $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ that assign to a Borel subset B of $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ the probability

$$\tilde{P}_{r,q}\{B\}\doteq\int_B\exp[2^r q\Phi(\mu)]P(W_{r,q}\in d\mu) \times\frac{1}{\int_{\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\exp[2^r q\Phi(\mu)]P(W_{r,q}\in d\mu)}. \tag{2.2}$$

From Corollary 2.5 we can easily derive the two-parameter Laplace principle for $\tilde{P}_{r,q}$ with rate function

$$J_\Phi(\mu)\doteq J(\mu)-\Phi(\mu)-\inf_{\nu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\{J(\nu)-\Phi(\nu)\}. \tag{2.3}$$

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

$$\begin{aligned}
&\lim_{r\to\infty}\lim_{q\to\infty}\frac{1}{2^r q}\log\int_{\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\exp[2^r q\, h(\mu)]\,\tilde{P}_{r,q}(d\mu)\\
&\quad=\lim_{r\to\infty}\lim_{q\to\infty}\frac{1}{2^r q}\log\int_{\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\exp\big[2^r q\,(h(\mu)+\Phi(\mu))\big]\,P(W_{r,q}\in d\mu)\\
&\qquad-\lim_{r\to\infty}\lim_{q\to\infty}\frac{1}{2^r q}\log\int_{\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\exp[2^r q\,\Phi(\mu)]\,P(W_{r,q}\in d\mu)\\
&\quad=\sup_{\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\big\{h(\mu)+\Phi(\mu)-J(\mu)\big\}-\sup_{\nu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\{\Phi(\nu)-J(\nu)\}\\
&\quad=\sup_{\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})}\{h(\mu)-J_\Phi(\mu)\}.
\end{aligned}$$

Since the two-parameter Laplace principle implies the two-parameter LDP with the same rate function, we have proved part (a) of the next corollary. Part (b) will enable us to characterize the equilibrium states of the continuum limit of the model of two-dimensional turbulence to be considered in [2].

COROLLARY 2.6. *Under the same conditions as in Theorem* 2.4, *the following conclusions hold*.

(a) *The sequence of measures* $\tilde{P}_{r,q}$ *defined in* (2.2) *satisfies the two-parameter Laplace principle and LDP on* $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ *with rate function* J_Φ *defined in* (2.3).

(b) *The set* $\mathscr{E}\doteq\{\mu\in\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})\colon J_\Phi(\mu)=0\}$ *is a nonempty compact subset of* $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$. *Furthermore*, *if* B *is a Borel subset of* $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ *whose closure has empty intersection with* $\mathscr{E}$, *then* $\lim_{r\to\infty}\lim_{q\to\infty}\tilde{P}_{r,q}\{B\}=0$.

The assertion about $\mathscr{E}$ in part (b) of the corollary follows from the fact that J_Φ is a rate function. The second assertion in part (b) is a consequence of the fact that if the closure of B, $\bar{B}$, has empty intersection with $\mathscr{E}$, then $J_\Phi(\bar{B})>0$. Hence by the large deviation upper bound in (a)

$$\tilde{P}_{r,q}\{B\}\le C\exp(-J_\Phi(\bar{B})/2)\to 0$$

as $q\to\infty$, $r\to\infty$.

The assertion in Theorem 2.4 that J is a rate function requires showing that J has compact level sets. This proof, given in Section 3.5, is surprisingly complicated, and it is only in this proof that part (iv) of Condition 2.1 is needed. The main effort is required to show that J is lower semicontinuous. This property, together with other estimates to be obtained, will yield the compactness of the level sets of J. The requirement that a function governing the large deviations of a process have compact level sets is not required in all aspects of the theory. For example, it is not needed to prove that an LDP implies a corresponding Laplace principle, although it is needed to show the reverse implication.

We next present four cases of processes L_q for which the LDP in Theorem 2.4 is of interest. Case (a) involves the empirical measures of i.i.d. random variables, case (b) the α-variate empirical measures of certain Markov chains for $\alpha \geq 2$ and case (d) the empirical processes of certain Markov chains. In all these cases, we can prove directly from the form of J that it has compact level sets, avoiding Condition 2.1 and the complicated proof in Section 3.5. However, in case (c), which involves the empirical measures of certain Markov chains, Condition 2.1 and the proof in Section 3.5 seem unavoidable since we cannot prove directly from the form of J that it has compact level sets.

EXAMPLE 2.7. (a) *Empirical measures of i.i.d. random variables.* Let $\mathscr{Y}$ be a Polish space and $\{\zeta_i, i \in \mathbb{N}\}$ a sequence of i.i.d. random variables taking values in $\mathscr{Y}$ and having common distribution ρ. We define $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ to be i.i.d. copies of

$$L_q \doteq \frac{1}{q} \sum_{i=1}^{q} \delta_{\zeta_i},$$

which takes values in $\mathscr{P}(\mathscr{Y})$. Sanov's theorem ([6], Theorem 2.2.1) implies that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with the convex rate function $I(\gamma) \doteq R(\gamma \mid \rho)$, where R is the relative entropy

$$R(\gamma \mid \rho) \doteq \begin{cases} \displaystyle\int_{\mathscr{Y}} \left(\log \frac{d\gamma}{d\rho}\right) d\gamma, & \text{if } \gamma \ll \rho, \\ \infty, & \text{otherwise.} \end{cases}$$

As a consequence of the chain rule ([6], Corollary C.3.2), if $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ has the decomposition $\mu = \theta \otimes \tau$, then

$$\begin{aligned} J(\mu) = J(\theta \otimes \tau) &= \int_{\mathscr{X}} R(\tau(x, \cdot) \mid \rho)\, \theta(dx) \\ &= R(\theta \otimes \tau \mid \theta \times \rho) = R(\mu \mid \theta \times \rho). \end{aligned}$$

Thus J has compact level sets since the relative entropy has this property ([6], Lemma 1.4.3(c)). This proof that J has compact level sets does not require Condition 2.1. According to Theorem 2.4, $W_{r,q}$ defined with these $L_{q,k}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function J.

(b) *α-variate empirical measures of certain Markov chains,* $\alpha \geq 2$. Let $\{\zeta_i, i \in \mathbb{N}\}$ be a Markov chain taking values in a Polish space $\mathscr{Y}$ and having transition probability function $p(y, dz)$. We denote by $\mathscr{C}_b(\mathscr{Y})$ the set of bounded continuous functions mapping $\mathscr{Y}$ into $\mathbb{R}$. It is assumed that $p(y, dz)$ satisfies the Feller property; that is, for all $f \in \mathscr{C}_b(\mathscr{Y})$ the function mapping

$$y \in \mathscr{Y} \mapsto (pf)(y) \doteq \int_{\mathscr{Y}} f(z)\, p(y, dz) \in \mathbb{R}$$

is continuous. It is also assumed that for some $C \in [1, \infty)$, all $y, y' \in \mathscr{Y}$, and all Borel subsets A of $\mathscr{Y}$

$$p(y, A) \leq C\, p(y', A). \tag{2.4}$$

We consider the bivariate empirical measures of the Markov chain, which is the case $\alpha = 2$ of the α-variate empirical measures. The case of general $\alpha \geq 3$ can be handled similarly and will be omitted. Let $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ be i.i.d. copies of the bivariate empirical measure

$$L_q \doteq \frac{1}{q}\sum_{i=1}^{q} \delta_{(\zeta_i,\, \zeta_{i+1})}, \tag{2.5}$$

which takes values in $\mathscr{P}(\mathscr{Y}^2)$. Given $\gamma \in \mathscr{P}(\mathscr{Y}^2)$, we denote by γ_1 and γ_2 the first and second marginals of γ obtained by projection onto the corresponding coordinates. Under the hypotheses on $p(y, dz)$, Theorem 1.4 in [9] proves that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y}^2)$ with the convex rate function

$$I(\gamma) \doteq \begin{cases} R(\gamma \mid \gamma_1 \otimes p), & \text{if } \gamma_1 = \gamma_2, \\ \infty, & \text{otherwise.} \end{cases} \tag{2.6}$$

Denote by $J^{(2)}$ the function on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^2)$ defined in Definition 2.3 in terms of this I. According to Theorem 2.4, $W_{r,q}$ defined with these $L_{q,k}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^2)$ with rate function $J^{(2)}$.

At the end of this section we prove directly from the form of $J^{(2)}$ that this function has compact level sets. Condition 2.1 is not required. In order to carry this out, it is useful to rewrite $J^{(2)}$. Given $\mu = \mu(dx \times dz_1 \times dz_2) \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^2)$, we denote by μ_i, $i = 1, 2, 3$, the ith marginal of μ obtained by projection onto the ith coordinate and by $\mu_{i,j}$, $1 \leq i < j \leq 3$, the marginal of μ obtained by projection onto the ith and jth coordinates. The measure μ_1 equals θ, and if $J^{(2)}(\mu) < \infty$, then it follows from (2.6) and the chain rule that $\mu_2 = \mu_3$ and

$$J^{(2)}(\mu) = R\big(\mu(dx \times dz_1 \times dz_2) \mid \mu_{1,2}(dx \times dz_1) \otimes p(z_1, dz_2)\big). \tag{2.7}$$

(c) *Empirical measures of certain Markov chains.* Let $\{\zeta_i,\ i \in \mathbb{N}\}$ be a Markov chain taking values in a Polish space $\mathscr{Y}$ and having transition probability function $p(y, dz)$. We assume that $p(y, dz)$ satisfies the same properties as in case (b). Let $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ be i.i.d. copies of the empirical measure

$$L_q \doteq \frac{1}{q}\sum_{i=1}^{q} \delta_{\zeta_i}.$$

Under the hypotheses on $p(y, dz)$, [21] proves that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with the convex rate function

$$I(\mu) \doteq \sup_{u \in \mathscr{U}(\mathscr{Y})} \int_{\mathscr{Y}} \frac{u(y)}{(pu)(y)}\, \mu(dy), \tag{2.8}$$

where $\mathscr{U}(\mathscr{Y})$ denotes the set of $u \in \mathscr{C}_b(\mathscr{Y})$ satisfying $u \geq \varepsilon$ on $\mathscr{Y}$ for some $\varepsilon = \varepsilon(u) > 0$. See also Chapter IV of [4]. Denote by J the function on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ defined in Definition 2.3 in terms of this I. According to Theorem 2.4, $W_{r,q}$ defined with these $L_{q,k}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function J. We cannot prove directly from the form of J that it has compact level sets, but must resort to the proof in Section 3.5 which requires Condition 2.1.

Using the contraction principle ([3], Theorem 4.2.1), we can obtain another representation for the rate function J. Indeed, denote by $W_{r,q}^{(2)}$ the process defined in (2.1), where we take $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ to be independent copies of the bivariate empirical measure L_q in (2.5). As pointed out in part (b), $W_{r,q}^{(2)}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^2)$ with rate function $J^{(2)}$. Now let Γ denote the continuous function mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^2)$ into $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ defined by $\Gamma(\mu) \doteq \mu_{1,2}$. Since $\Gamma(W_{r,q}^{(2)}) = W_{r,q}$, it follows that $W_{r,q}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function

$$\tilde{J}(\gamma) \doteq \inf\left\{J^{(2)}(\mu) \colon \mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^2), \mu_{1,2} = \gamma\right\}.$$

Since a rate function for $W_{r,q}$ is unique, it follows that under Condition 2.1 J equals $\tilde{J}$.

(d) *Empirical processes of certain Markov chains.* Let $\{\zeta_i,\ i \in \mathbb{N}\}$ be a Markov chain taking values in a Polish space $\mathscr{Y}$ and having transition probability function $p(y, dz)$. We assume that $p(y, dz)$ satisfies the same properties as in case (b). Define $\mathscr{Y}^{\mathbb{Z}}$ to be the product space $\prod_{j\in\mathbb{Z}} \mathscr{Y}_j$, where for each j, $\mathscr{Y}_j \doteq \mathscr{Y}$, and let $\mathscr{P}_T(\mathscr{Y}^{\mathbb{Z}})$ denote the space of probability measures P on $\mathscr{Y}^{\mathbb{Z}}$ which satisfy $P \circ T = P$; T denotes the shift operator on $\mathscr{Y}^{\mathbb{Z}}$. For each $q \in \mathbb{N}$ we repeat the sequence $(\zeta_1, \zeta_2, \ldots, \zeta_q)$ periodically into a doubly infinite sequence, obtaining a point $\zeta(q) \in \mathscr{Y}^{\mathbb{Z}}$. We then let $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ be i.i.d. copies of the empirical process

$$L_q \doteq \frac{1}{q} \sum_{i=0}^{q-1} \delta_{T^i \zeta(q)},$$

which takes values in $\mathscr{P}_T(\mathscr{Y}^{\mathbb{Z}})$.

In order to specify the rate function, additional notation is needed. For $\vec{z} \doteq (\ldots, z_{-2}, z_{-1}, z_0, z_1, \ldots) \in \mathscr{Y}^{\mathbb{Z}}$, let $z_{-1}^{-\infty} \doteq (\ldots, z_{-2}, z_{-1})$. We denote by $\hat{X}_j$, $j \in \mathbb{Z}$, the mapping that takes $\vec{z}$ to z_j and by $\hat{X}_{-1}^{-\infty}$ the mapping that takes $\vec{z}$ to $z_{-1}^{-\infty}$; thus $\hat{X}_{-1}^{-\infty} = (\ldots, \hat{X}_{-2}, \hat{X}_{-1})$. For $P \in \mathscr{P}_T(\mathscr{Y}^{\mathbb{Z}})$ we define $P^*(z_{-1}^{-\infty}, dz_0)$ to be a regular conditional distribution, with respect to P, of $\hat{X}_0$ given $\hat{X}_{-1}^{-\infty} = z_{-1}^{-\infty}$, and we write $P^*(dz_{-1}^{-\infty})$ for the P-distribution of $\hat{X}_{-1}^{-\infty}$. Finally, let $\mathscr{Y}_{-1}^{-\infty} \doteq \prod_{j=-1}^{-\infty} \mathscr{Y}_j$.

Under the hypotheses on $p(y, dz)$, Theorem 1.3 in [11] proves that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y}^{\mathbb{Z}})$ with the convex rate function

$$(2.9)\quad I(P) \doteq \begin{cases} \displaystyle\int_{\mathscr{Y}_{-1}^{-\infty}} R\big(P^*(z_{-1}^{-\infty}, \cdot) \mid p(z_{-1}, \cdot)\big) P(dz_{-1}^{-\infty}), & \text{if } P \in \mathscr{P}_T(\mathscr{Y}^{\mathbb{Z}}), \\ \infty, & \text{otherwise.} \end{cases}$$

Denote by J the function on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}^{\mathbb{Z}})$ defined in Definition 2.3 in terms of this I. According to Theorem 2.4, $W_{r,q}$ defined with these $L_{q,k}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Z}^{\mathbb{Z}})$ with rate function J. While it is possible to prove directly from the form of J that it has compact level sets, this proof is omitted.

We end this section by giving a direct proof that the function $J^{(2)}$ in case (b) of Example 2.7 has compact level sets. This proof does not require Condition 2.1. For $M \in [0, \infty)$, we show that any sequence $\{\mu^n, n \in \mathbb{N}\}$ in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Z}^2)$ satisfying $J^{(2)}(\mu^n) \le M$ is tight; $J^{(2)}$ is given by (2.7). Since R is lower semicontinuous in both variables ([6], Lemma 1.4.3(b)) and p satisfies the Feller property, the proof that $J^{(2)}$ has compact level sets is completed by Prohorov's theorem.

We fix a point $y' \in \mathscr{Z}$ and set $\nu(dz_2) \doteq p(y', dz_2)$. By (2.7) and the Donsker–Varadhan variational formula ([6], Lemma 1.4.3(a)), for any $n \in \mathbb{N}$ and any $g \in \mathscr{C}_b(\mathscr{Z})$

$$\begin{aligned} M &\ge J^{(2)}(\mu^n) \\ &\ge \int_{\mathscr{X} \times \mathscr{Z}^2} g(z_2)\, \mu^n(dx \times dz_1 \times dz_2) \\ &\quad - \log \int_{\mathscr{X} \times \mathscr{Z}^2} \exp(g(z_2))(\mu^n)_{1,2}(dx \times dz_1) \otimes p(z_1, dz_2), \end{aligned}$$

and so by (2.4)

$$M + \log C \ge \int_{\mathscr{Z}} g(z_2)(\mu^n)_3(dz_2) - \log \int_{\mathscr{Z}} \exp(g(z_2))\nu(dz_2).$$

Since $g \in \mathscr{C}_b(\mathscr{Z})$ is arbitrary, we conclude that for any $n \in \mathbb{N}$

$$M + \log C \ge R\big((\mu^n)_3 \,|\, \nu\big).$$

The compactness of the level sets of $R(\cdot \,|\, \nu)$ implies that the sequence of marginals $\{(\mu^n)_3\}$ is tight, and since $J^{(2)}(\mu^n) \le M < \infty$, the sequence $\{(\mu^n)_2\} = \{(\mu^n)_3\}$ is also tight. Since $(\mu^n)_1 = \theta$ is tight, we conclude that $\{\mu^n, n \in \mathbb{N}\}$ is tight. This completes the proof that $J^{(2)}$ has compact level sets.

3. Proof of Theorem 2.4. This section consists of five subsections, which prove the following: (1) a number of preliminary lemmas, (2) the large deviation lower bound in Theorem 2.4, (3) the exponential tightness of $\{W_{r,q}, q \in \mathbb{N}\}$ for each fixed r, (4) the large deviation upper bound in Theorem 2.4, and (5) the compactness of the level sets of J. Together, (2), (4) and (5) yield the theorem.

We metrize $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ with the dual-bounded-Lipschitz metric defined in [5], page 310. $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$ is a Polish space with respect to this metric, which is compatible with the topology of weak convergence ([5], Property 11.3.2, Theorem 11.3.3, Corollary 11.5.5). As a closed subset of $\mathscr{P}(\mathscr{X} \times \mathscr{Y})$,

$$\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}) \doteq \big\{\mu \in \mathscr{P}(\mathscr{X} \times \mathscr{Y}) \colon \mu_1 = \theta\big\}$$

is also a Polish space when metrized by d. Given $\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ and $\varepsilon > 0$, $B(\mu, \varepsilon)$ 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. In the present paper, the form of the metric d is not used; for example, the Lévy–Prohorov metric could be employed in its place. However, in the application to two-dimensional turbulence given in [2], the use of the metric d facilitates the proof of a key estimate.

3.1. *A number of preliminary lemmas.* The second lemma in this subsection is an approximation result for stochastic kernels needed in the proof of the large deviation lower bound and in the proof that J has compact level sets. In order to prove this lemma, we first need to know that $\sigma(\bigcup_{r\in\mathbb{N}}\mathscr{F}_r) = \mathscr{B}(\mathscr{X})$, where $\mathscr{F}_r$ denotes the σ-field generated by the partition Ξ_r and $\mathscr{B}(\mathscr{X})$ denotes the Borel σ-field of $\mathscr{X}$.

LEMMA 3.1. *Let $\Xi_r \doteq \{D_{r,k}, k = 1, \ldots, 2^r\}$ be a partition of $\mathscr{X}$ satisfying parts* (ii) *and* (iii) *of Condition* 2.1. *Then $\sigma(\bigcup_{r\in\mathbb{N}}\mathscr{F}_r) = \mathscr{B}(\mathscr{X})$.*

PROOF. Let F be any closed subset of $\mathscr{X}$. It suffices to prove that there exists a sequence $\{D_r(F), r \in \mathbb{N}\}$ of subsets of $\mathscr{X}$ such that $D_r(F) \in \mathscr{F}_r$ for each r and $D_r(F) \downarrow F$. Define

$$A_r \doteq \big\{k \in \{1, \ldots, 2^r\}: D_{r,k} \cap F \neq \emptyset\big\} \quad \text{and} \quad D_r(F) \doteq \bigcup_{k\in A_r} D_{r,k}.$$

For each $r \in \mathbb{N}$, $D_r(F) \in \mathscr{F}_r$, and due to part (ii) of Condition 2.1, $D_{r+1}(F) \subset D_r(F)$. It must be shown that $\bigcap_{r=1}^{\infty} D_r(F) = F$. Since $F \subset D_r(F)$ for each r, one direction of containment is obvious. We prove by contradiction that $\bigcap_{r=1}^{\infty} D_r(F) \subset F$. Thus suppose that there exists $x \in \bigcap_{r=1}^{\infty} D_r(F)$ and $x \notin F$. Since F is closed, there exists $\delta > 0$ such that $m(x, F) > \delta$, where m denotes the metric on $\mathscr{X}$. For all $r \in \mathbb{N}$, $x \in D_r(F)$ and so $x \in D_{r,k}$ for some $k \in A_r$. Hence there exists $y_r \in D_{r,k} \cap F$. It follows that for every $r \in \mathbb{N}$

$$0 < \delta < m(x, F) \le m(x, y_r) \le \operatorname{diam}(D_{r,k}) \le \max_{i=1,\ldots,2^r} \operatorname{diam}(D_{r,i}).$$

By part (iii) of Condition 2.1, the right-hand side of this display tends to zero as $r \to \infty$. This contradiction completes the proof. □

The next lemma approximates an arbitrary stochastic kernel by a stochastic kernel which, when viewed as a mapping from $\mathscr{X}$ into $\mathscr{P}(\mathscr{Y})$, is constant on the cells of the partition Ξ_r.

LEMMA 3.2. *We assume parts* (i)–(iii) *of Condition* 2.1. *Let μ be any measure in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with the decomposition $\mu = \theta \otimes \tau$, where $\tau(x, dy)$ is a stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$. For $r \in \mathbb{N}$ and $k \in \{1, \ldots, 2^r\}$ define the probability measures on $\mathscr{Y}$*

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

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

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

Then as $r \to \infty$, $\tau^r(x, \cdot) \Rightarrow \tau(x, \cdot)$ θ-a.s. for $x \in \mathscr{X}$, and $\theta \otimes \tau^r \Rightarrow \theta \otimes \tau = \mu$.

PROOF. Let f be any bounded continuous function mapping $\mathscr{Y}$ into $\mathbb{R}$. Suppose we can prove that θ-a.s. for $x \in \mathscr{X}$

$$\lim_{r\to\infty} \int_{\mathscr{Y}} f(y)\, \tau^r(x, dy) = \int_{\mathscr{Y}} f(y)\, \tau(x, dy). \tag{3.1}$$

Then a standard separability argument implies that $\tau^r(x, \cdot) \Rightarrow \tau(x, \cdot)$ θ-a.s. for $x \in \mathscr{X}$. From this one can show the weak convergence $\theta \otimes \tau^r \Rightarrow \theta \otimes \tau$ by considering integrals with respect to $g(x)h(y)$ for $g \in \mathscr{C}_b(\mathscr{X})$ and $h \in \mathscr{C}_b(\mathscr{Y})$ ([12], Proposition 4.6, page 115). The Lebesgue dominated convergence theorem completes the proof.

We now prove (3.1). For $r \in \mathbb{N}$ denote by $\mathscr{F}_r$ the σ-field generated by the partition Ξ_r and define

$$X_r^f(x) \doteq \int_{\mathscr{Y}} f(y)\, \tau^r(x, dy) \quad \text{and} \quad X^f(x) \doteq \int_{\mathscr{Y}} f(y)\, \tau(x, dy).$$

X_r^f is an $\mathscr{F}_r$-measurable function mapping $\mathscr{X}$ into $\mathbb{R}$ and $\sup_{r\in\mathbb{N}} E_\theta|X_r^f| \le \|f\|_\infty$. By part (ii) of Condition 2.1 $\mathscr{F}_r \subset \mathscr{F}_{r+1}$, and for any $B \in \mathscr{F}_r$

$$\int_B X_r^f\, d\theta = \int_B X^f\, d\theta.$$

Thus $X_r^f = E\{X^f \mid \mathscr{F}_r\}$. A standard result of Lévy ([20], Theorem 3, page 510) and Lemma 3.1 yield the θ-a.s. convergence

$$\lim_{r\to\infty} X_r^f = \lim_{r\to\infty} E\{X^f \mid \mathscr{F}_r\} = E\left\{X^f \,\middle|\, \sigma\left(\bigcup_{r\in\mathbb{N}} \mathscr{F}_r\right)\right\} = X^f.$$

This completes the proof of the lemma. □

The following lemma is a consequence of Lemmas 2.5, 2.7, 2.8 in [14] and the assumptions that $L_{q,1}, \ldots, L_{q,2^r}$ are i.i.d. copies of L_q and that the sequence $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with rate function I.

LEMMA 3.3. *For each r the sequence $\{(L_{q,1}, \ldots, L_{q,2^r}), q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})^{2^r}$ with rate function*

$$(\nu_1, \ldots, \nu_{2^r}) \mapsto \sum_{k=1}^{2^r} I(\nu_k).$$

The next lemma shows that the convex rate function I on $\mathscr{P}(\mathscr{Y})$ can be represented in terms of a Legendre–Fenchel transform. In the case that $I(\cdot)$ equals the relative entropy $R(\cdot\,|\,\rho)$ for some $\rho \in \mathscr{P}(\mathscr{Y})$, this lemma states the Donsker–Varadhan variational formula for the relative entropy. In this case, for $f \in \mathscr{C}_b(\mathscr{Y})$, $I^*(f) = \log \int_{\mathscr{Y}} \exp(f)\, d\rho$.

LEMMA 3.4. *Let $\mathscr{Z}$ be a Polish space and Ψ a convex, lower semicontinuous function mapping $\mathscr{P}(\mathscr{Z})$ into $(-\infty, \infty]$ such that Ψ is not identically equal to ∞; in particular, let Ψ equal the convex rate function I on $\mathscr{P}(\mathscr{Y})$. Define $\Psi^*\colon \mathscr{C}_b(\mathscr{Z}) \mapsto (-\infty, \infty]$ by*

$$\Psi^*(f) \doteq \sup_{\nu \in \mathscr{P}(\mathscr{Z})} \left\{ \int_{\mathscr{Z}} f\, d\nu - \Psi(\nu) \right\}.$$

Then for all $\nu \in \mathscr{P}(\mathscr{Z})$

$$\Psi(\nu) = \sup_{f \in \mathscr{C}_b(\mathscr{Z})} \left\{ \int_{\mathscr{Z}} f\, d\nu - \Psi^*(f) \right\}.$$

PROOF. Let $\mathscr{M}(\mathscr{Z})$ denote the space of finite signed measures on $\mathscr{Z}$. With the topology of weak convergence, $\mathscr{M}(\mathscr{Z})$ is a locally convex, Hausdorff topological space whose topological dual is $\mathscr{C}_b(\mathscr{Z})$ ([4], Lemma 3.2.3). We extend Ψ to a convex, lower semicontinuous function $\tilde{\Psi}$ on $\mathscr{M}(\mathscr{Z})$ by defining $\tilde{\Psi}(\nu) \doteq \Psi(\nu)$ for $\nu \in \mathscr{P}(\mathscr{Z})$ and $\tilde{\Psi}(\nu) \doteq \infty$ for $\nu \in \mathscr{M}(\mathscr{Z}) \setminus \mathscr{P}(\mathscr{Z})$. Then for any $f \in \mathscr{C}_b(\mathscr{Z})$

$$\tilde{\Psi}^*(f) \doteq \sup_{\nu \in \mathscr{M}(\mathscr{Z})} \left\{ \int_{\mathscr{Z}} f\, d\nu - \tilde{\Psi}(\nu) \right\} = \Psi^*(f),$$

and thus by ([4], Theorem 2.2.15), for any $\nu \in \mathscr{M}(\mathscr{Z})$

$$\tilde{\Psi}(\nu) = \sup_{f \in \mathscr{C}_b(\mathscr{Z})} \left\{ \int_{\mathscr{Z}} f\, d\nu - \tilde{\Psi}^*(f) \right\} = \sup_{f \in \mathscr{C}_b(\mathscr{Z})} \left\{ \int_{\mathscr{Z}} f\, d\nu - \Psi^*(f) \right\}.$$

Taking $\nu \in \mathscr{P}(\mathscr{Z})$ completes the proof. $\square$

We now prove a Jensen-type inequality that involves the convex rate function I on $\mathscr{P}(\mathscr{Y})$. This is needed in the proof of the large deviation lower bound and in the proof that J has compact level sets.

LEMMA 3.5. *Let γ be a probability measure on $\mathscr{X}$ and τ a stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$. Then*

$$\int_{\mathscr{X}} I(\tau(x, \cdot))\gamma(dx) \geq I\left(\int_{\mathscr{X}} \tau(x, \cdot)\gamma(dx) \right).$$

REMARK 3.6. In the case that $I(\cdot) = R(\cdot\,|\,\rho)$ for some $\rho \in \mathscr{P}(\mathscr{Y})$, the lemma states the well-known fact that $R(\mu\,|\,\gamma \times \rho) \geq R(\mu_2\,|\,\rho)$. This follows immediately from the chain rule and the Donsker–Varadhan variational formula for the relative entropy.

PROOF. By Lemma 3.4, for any $f \in \mathscr{C}_b(\mathscr{Y})$

$$\int_{\mathscr{X}} I(\tau(x,\cdot))\,\theta(dx) \geq \int_{\mathscr{X}} \left[\int_{\mathscr{Y}} f(y)\tau(x,dy) - I^*(f)\right]\theta(dx)$$
$$= \int_{\mathscr{Y}} f(y)\,d\gamma(y) - I^*(f),$$

where $\gamma(\cdot) \doteq \int_{\mathscr{X}} \tau(x,\cdot)\theta(dx)$. Taking the supremum over $f \in \mathscr{C}_b(\mathscr{Y})$ gives

$$\int_{\mathscr{X}} I(\tau(x,\cdot))\theta(dx) \geq I(\gamma) = I\left(\int_{\mathscr{X}} \tau(x,\cdot)\,\theta(dx)\right).$$

This completes the proof. □

3.2. *Proof of the large deviation lower bound.* Under parts (i)–(iii) of Condition 2.1 we prove that for any open subset G of $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$

$$\liminf_{r\to\infty}\liminf_{q\to\infty}\frac{1}{2^r q}\log P\{W_{r,q}\in G\} \geq -J(G). \tag{3.2}$$

Let $\mu = \theta\otimes\tau$ be any measure in G and choose $\varepsilon > 0$ so that $B(\theta\otimes\tau,\varepsilon)\subset G$. Also choose $N\in\mathbb{N}$ such that for all $r\geq N$ $B(\theta\otimes\tau^r,\varepsilon/2)\subset B(\theta\otimes\tau,\varepsilon)$, where τ^r is the stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$ defined in Lemma 3.2 in terms of τ. Such an N exists because of the weak convergence $\theta\otimes\tau^r \Rightarrow \theta\otimes\tau$ proved in that lemma. Finally, define the open set

$$G_{r,\varepsilon} \doteq \left\{(\nu_1,\ldots,\nu_{2^r})\in\mathscr{P}(\mathscr{Y})^{2^r}\colon \theta(dx)\otimes\sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\nu_k(dy) \in B(\theta\otimes\tau^r,\varepsilon/2)\right\}. \tag{3.3}$$

Then for all $r\geq N$ Lemmas 3.3 and 3.5 yield

$$\begin{aligned}
&\liminf_{q\to\infty}\frac{1}{2^r q}\log P\{W_{r,q}\in G\}\\
&\quad\geq \liminf_{q\to\infty}\frac{1}{2^r q}\log P\left\{W_{r,q}\in B(\theta\otimes\tau^r,\varepsilon/2)\right\}\\
&\quad= \liminf_{q\to\infty}\frac{1}{2^r q}\log P\{(L_{q,1},\ldots,L_{q,2^r})\in G_{r,\varepsilon}\}\\
&\quad\geq -\frac{1}{2^r}\inf\left\{\sum_{k=1}^{2^r} I(\nu_k)\colon (\nu_1,\ldots,\nu_{2^r})\in G_{r,\varepsilon}\right\}\\
&\quad\geq -\frac{1}{2^r}\sum_{k=1}^{2^r} I(\tau_k^r)\\
&\quad= -\frac{1}{2^r}\sum_{k=1}^{2^r} I\left(2^r\int_{D_{r,k}}\tau(x,\cdot)\theta(dx)\right)
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
&\geq -\sum_{k=1}^{2^r} \int_{D_{r,k}} I(\tau(x,\cdot))\theta(dx) \\
&= -\int_{\mathscr{X}} I(\tau(x,\cdot))\,\theta(dx) \\
&= -J(\mu).
\end{aligned}$$

This implies that

$$\liminf_{r\to\infty} \liminf_{q\to\infty} \frac{1}{2^r q} \log P\{W_{r,q} \in G\} \geq -J(\mu).$$

Since $\mu \in G$ is arbitrary, the proof of the lower bound (3.2) is complete. □

For use in Section 3.5, we record the content of display (3.4).

COROLLARY 3.7. *For any open subset G of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ and $\mu \in G$, there exists $N \in \mathbb{N}$ such that for all $r \geq N$*

$$\liminf_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in G\} \geq -2^r J(\mu).$$

3.3. *Proof of exponential tightness.* We prove that for each fixed $r \in \mathbb{N}$ the sequence $\{W_{r,q}, q \in \mathbb{N}\}$ is exponentially tight. For this proof Condition 2.1 is not needed. The exponential tightness will be used in Section 3.4 to pass from the large deviation upper bound for fixed r and for compact subsets of $\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})$ to the large deviation upper bound for fixed r and for closed subsets of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$.

LEMMA 3.8. *Fix $r \in \mathbb{N}$. Then for each $M \in (0, \infty)$ there exists a compact subset K_M of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ such that*

$$\limsup_{q\to\infty} \frac{1}{q} \log P\big\{W_{r,q} \in (K_M)^c\big\} \leq -M.$$

PROOF. Since $L_{q,1}, \ldots, L_{q,2^r}$ have the same distributions as L_q and $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$, there is a compact subset D_M of $\mathscr{P}(\mathscr{Y})$ such that for all $k \in \{1, \ldots, 2^r\}$ ([14], Lemma 2.6)

$$\limsup_{q\to\infty} \frac{1}{q} \log P\big\{L_{q,k} \in (D_M)^c\big\} \leq -M.$$

Prohorov's theorem implies that for each $n \in \mathbb{N}$ there exists a compact subset A_n of $\mathscr{Y}$ such that

$$\inf_{\sigma\in D_M} \sigma\{A_n\} \geq 1 - \frac{1}{n}.$$

Hence

$$D_M \subset \Delta_M \doteq \bigcap_{n=1}^{\infty} \left\{\sigma \in \mathscr{P}(\mathscr{Y})\colon \sigma\{A_n\} \geq 1 - \frac{1}{n}\right\}.$$

Since Δ_M is a compact subset of $\mathscr{P}(\mathscr{Y})$,

$$\tilde{\Delta}_M \doteq \{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})\colon \mu_2 \in \Delta_M\}$$

is a compact subset of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. The convexity of Δ_M now allows us to write

$$\begin{aligned}
&\limsup_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in (\tilde{\Delta}_M)^c\} \\
&\quad = \limsup_{q\to\infty} \frac{1}{q} \log P\left\{\frac{1}{2^r} \sum_{k=1}^{2^r} L_{q,k} \in (\Delta_M)^c\right\} \\
&\quad \le \limsup_{q\to\infty} \frac{1}{q} \log P\left\{\bigcup_{k=1}^{2^r} \{L_{q,k} \in (\Delta_M)^c\}\right\} \\
&\quad = \max_{k=1,\dots,2^r} \limsup_{q\to\infty} \frac{1}{q} \log P\{L_{q,k} \in (\Delta_M)^c\} \\
&\quad \le \max_{k=1,\dots,2^r} \limsup_{q\to\infty} \frac{1}{q} \log P\{L_{q,k} \in (D_M)^c\} \\
&\quad \le -M
\end{aligned}$$

This yields the lemma with $K_M \doteq \tilde{\Delta}_M$. $\square$

3.4. *Proof of the large deviation upper bound.* The large deviation upper bound states that for any closed subset F of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

$$\limsup_{r\to\infty} \limsup_{q\to\infty} \frac{1}{2^r q} \log P\{W_{r,q} \in F\} \le -J(F). \tag{3.5}$$

The upper bound is first derived for compact subsets of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, for which it suffices to prove the bound for an arbitrary closed ball in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. As we will see, the exponential tightness proved in Lemma 3.8 will then yield (3.5) for all closed subsets F of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. Throughout this section we assume only part (i) of Condition 2.1.

For any $\mu = \theta \otimes \tau \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, $\varepsilon > 0$, and $r \in \mathbb{N}$, we define the closed set

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

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

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

for some $\nu_1, \dots, \nu_{2^r} \in \mathscr{P}(\mathscr{Y})$. By Lemma 3.3, since $\theta\{D_{r,k}\} = 1/2^r$,

$$\begin{aligned}
&\limsup_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in \bar{B}(\mu, \varepsilon)\} \\
&\quad = \limsup_{q\to\infty} \frac{1}{q} \log P\}(L_{q,1}, \dots, L_{q,2^r}) \in F_{r,\varepsilon}\} \\
&\quad \le -2^r \inf\left\{ \frac{1}{2^r} \sum_{k=1}^{2^r} I(\nu_k) \colon (\nu_1, \dots, \nu_{2^r}) \in F_{r,\varepsilon} \right\} \\
&\quad = -2^r \inf\left\{ \int_{\mathscr{X}} I(\nu(x,\cdot))\, \theta(dx) \colon \theta(dx) \otimes \nu(x, dy) \in \bar{B}(\mu, \varepsilon) \cap \mathscr{P}_{\theta, r}(\mathscr{X} \times \mathscr{Y}) \right\} \\
&\quad \le -2^r \inf\left\{ \int_{\mathscr{X}} I(\nu(x,\cdot))\theta(dx) \colon \theta(dx) \otimes \nu(x, dy) \in \bar{B}(\mu, \varepsilon) \right\} \\
&\quad = -2^r J(\bar{B}(\mu, \varepsilon)).
\end{aligned}$$

Since $\bar{B}(\mu, \varepsilon)$ is an arbitrary closed ball in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, the last display implies that for any $r \in \mathbb{N}$ and any compact subset K of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$

$$\limsup_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in K\} \le -2^r J(K). \tag{3.6}$$

Thus, by the exponential tightness proved in Lemma 3.8, for all $r \in \mathbb{N}$ and all closed subsets F of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ ([4], Lemma 2.1.5)

$$\limsup_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in F\} \le -2^r J(F).$$

Dividing both sides of this inequality by 2^r and taking the limit superior as $r \to \infty$ yield the desired upper bound (3.5). □

3.5. *Proof of compact level sets of J.* The main effort in this subsection will be to prove that under Condition 2.1,

$$J(\mu) \doteq \int_{\mathscr{X}} I(\tau(x,\cdot))\, \theta(dx)$$

is lower semicontinuous. By adapting Lemma 2.1.5 in [4], we now show that the lower semicontinuity of J implies that J has compact level sets. According to Lemma 3.8, for each $r \in \mathbb{N}$ and $M \in [0, \infty)$ there exists a compact subset $K_{2^r(M+1)}$ in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ such that

$$\limsup_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in (K_{2^r(M+1)})^c\} \le -2^r(M+1).$$

Given $\delta \in (0, 1)$, choose $\mu \in (K_{2^r(M+1)})^c$ such that

$$J(\mu) \le J((K_{2^r(M+1)})^c) + \delta.$$

Then by Corollary 3.7 there exists $N \in \mathbb{N}$ such that for all $r \geq N$

$$\liminf_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in (K_{2^r(M+1)})^c\} \geq -2^r J(\mu) \geq -2^r J\big((K_{2^r(M+1)})^c\big) - 2^r \delta.$$

Hence for any $r \geq N$

$$\begin{aligned} -2^r(M+1) &\geq \limsup_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in (K_{2^r(M+1)})^c\} \\ &\geq \liminf_{q\to\infty} \frac{1}{q} \log P\{W_{r,q} \in (K_{2^r(M+1)})^c\} \\ &\geq -2^r J\big((K_{2^r(M+1)})^c\big) - 2^r \delta. \end{aligned}$$

It follows that $J((K_{2^r(M+1)})^c) \geq M + 1 - \delta > M$ or that

$$\big\{\mu \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y}) \colon J(\mu) \leq M\big\} \subset K_{2^r(M+1)}.$$

The lower semicontinuity of J and the compactness of $K_{2^r(M+1)}$ imply that $\{\mu \colon J(\mu) \leq M\}$ is compact.

The proof that J is lower semicontinuous uses the next weak convergence result.

LEMMA 3.9. *Let $\{\gamma^n, n \in \mathbb{N}\}$ be a sequence of probability measures on a Polish space $\mathscr{Z}$ that converges weakly to γ and let B be a Borel subset of $\mathscr{Z}$ satisfying $\gamma\{\partial B\} = 0$. Then as subprobability measures on $\mathscr{Z}$ $\gamma^n\{B \cap \cdot\} \Rightarrow \gamma\{B \cap \cdot\}$.*

PROOF. By the portmanteau theorem it suffices to prove that for any closed subset F of $\mathscr{Z}$

$$\limsup_{n\to\infty} \gamma^n\{B \cap F\} \leq \gamma\{B \cap F\}.$$

We denote the metric on $\mathscr{Z}$ by m. As pointed out on page 14 of [1], there exists a sequence $\delta_k \to 0$ such that each set $F_k \doteq \{x \in \mathscr{Z} \colon m(x, F) \leq \delta_k\}$ satisfies $\gamma\{\partial F_k\} = 0$. Since

$$\gamma\{\partial(B \cap F_k)\} \leq \gamma\{\partial B\} + \gamma\{\partial F_k\} = 0$$

and $\gamma^n \Rightarrow \gamma$,

$$\begin{aligned} \limsup_{n\to\infty} \gamma^n\{B \cap F\} &\leq \limsup_{n\to\infty} \gamma^n\{B \cap F_k\} = \lim_{n\to\infty} \gamma^n\{B \cap F_k\} \\ &= \gamma\{B \cap F_k\}. \end{aligned}$$

Sending $k \to \infty$ completes the proof of the lemma since $B \cap F_k \downarrow B \cap F$ and thus $\gamma\{B \cap F_k\} \downarrow \gamma\{B \cap F\}$. $\square$

Proof that J is lower semicontinuous. We prove the lower semicontinuity via Lemma 3.9 and the following lemma, which produces a sequence $\{J_r, r \in \mathbb{N}\}$ of functions mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$ with three properties. Properties (a) and (b) imply that $J = \sup_{r\in\mathbb{N}} J_r$. The lower semicontinuity of each J_r proved in (c) yields the lower semicontinuity of J. Part (iv) of Condition 2.1 is used to prove property (c).

LEMMA 3.10. *There exists a sequence* $\{J_r, r \in \mathbb{N}\}$ *of functions mapping* $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ *into* $\mathbb{R}$ *with the following three properties*:

(a) $J_r \le J$;
(b) $\liminf_{r\to\infty} J_r \ge J$;
(c) J_r *is lower semicontinuous*.

PROOF. For each $r \in \mathbb{N}$, $k \in \{1, \dots, 2^r\}$, and $\mu = \theta \otimes \tau \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$, let $\theta_{r,k}(dx)$ denote the probability measure $2^r 1_{D_{r,k}}(x)\theta(dx)$ on $\mathscr{X}$ and let $\tau^r(x, dy)$ be the stochastic kernel defined in Lemma 3.2 in terms of τ. We define

$$J_r(\mu) \doteq \int_{\mathscr{X}} I(\tau^r(x,\cdot))\,\theta(dx) = \frac{1}{2^r}\sum_{k=1}^{2^r} I\biggl(\int_{\mathscr{X}} \tau(x,\cdot)\theta_{r,k}(dx)\biggr).$$

We prove that J_r has the three properties given in the statement of Lemma 3.10.

Property (a). This is a consequence of Lemma 3.5, which yields

$$J_r(\mu) = \frac{1}{2^r}\sum_{k=1}^{2^r} I\biggl(\int_{\mathscr{X}} \tau(x,\cdot)\,\theta_{r,k}(dx)\biggr) \le \frac{1}{2^r}\sum_{k=1}^{2^r}\int_{\mathscr{X}} I(\tau(x,\cdot))\,\theta_{r,k}(dx) = J(\mu).$$

Property (b). Since $\tau^r(x,\cdot) \Rightarrow \tau(x,\cdot)$ θ-a.s. for $x \in \mathscr{X}$ (Lemma 3.2), Fatou's lemma and the nonnegativity and lower semicontinuity of I imply

$$\liminf_{r\to\infty} J_r(\mu) = \liminf_{r\to\infty}\int_{\mathscr{X}} I\bigl(\tau^r(x,\cdot)\bigr)\theta(dx) \ge \int_{\mathscr{X}} I(\tau(x,\cdot))\theta(dx) = J(\mu).$$

Property (c). We must show that if $\mu^n = \theta\otimes\tau^n \in \mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y}) \Rightarrow \mu = \theta\otimes\tau$, then $\liminf_{n\to\infty} J_r(\mu^n) \ge J_r(\mu)$. Since I is nonnegative and lower semicontinuous, this reduces to showing for all $k \in \{1,\dots,2^r\}$ that as probability measures on $\mathscr{Y}$

$$\begin{aligned}\int_{\mathscr{X}} \tau^n(x,\cdot)\theta_{r,k}(dx) &= 2^r\mu^n\{D_{r,k}\times\cdot\} = 2^r\mu^n\bigl\{(D_{r,k}\times\mathscr{Y})\cap(\mathscr{X}\times\cdot)\bigr\}\\ &\Rightarrow 2^r\mu\bigl\{(D_{r,k}\times\mathscr{Y})\cap(\mathscr{X}\times\cdot)\bigr\}\\ &= 2^r\mu\{D_{r,k}\times\cdot\} = \int_{\mathscr{X}}\tau(x,\cdot)\,\theta_{k,r}(dx).\end{aligned}$$

Since by part (iv) of Condition 2.1 $\mu\{\partial(D_{r,k}\times\mathscr{Y})\} = \theta\{\partial D_{r,k}\} = 0$, the weak convergence in the last display is a consequence of Lemma 3.9. □

With this proof that J has compact level sets, the proof of Theorem 2.4 is complete.

4. Restatement of results in Section 2. Theorem 2.4 and Corollary 2.5 state the two-parameter LDP and Laplace principle for the sequence of random probability measures

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

which take values in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. In those results it is assumed that $L_{q,1}, L_{q,2}, \ldots, L_{q,2^r}$ are i.i.d. copies of a random measure L_q on $\mathscr{Y}$ and that $\{L_q, q \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with a convex rate function I. In preparation for applications to the statistical mechanics of turbulence [2], we now restate those results for a class of processes defined like $W_{r,q}$ but with $L_{q,1}, \ldots, L_{q,2^r}$ replaced by other sequences of random measures which satisfy the LDP with different scaling constants.

These applications all involve a sequence of random measures $\{Y_n, n \in \mathbb{N}\}$ for which the LDP will be proved in a novel way; namely, by approximating it by a suitable doubly indexed sequence for which the LDP is readily available. The doubly indexed sequences that typically arise are defined next. Let $(\Omega, \mathscr{F}, P)$ be a probability space, $\mathscr{Y}$ a Polish space and I a convex rate function on $\mathscr{P}(\mathscr{Y})$. For each n and r in $\mathbb{N}$, we consider a sequence $L_{n,r,1}, L_{n,r,2}, \ldots, L_{n,r,2^r}$ of independent random variables mapping Ω into $\mathscr{P}(\mathscr{Y})$ with the property that for each $r \in \mathbb{N}$ and $k = 1, 2, \ldots, 2^r$, $\{L_{n,r,k}, n \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with scaling constants $n/2^r$ and rate function I. In other words, for any closed subset F of $\mathscr{P}(\mathscr{Y})$

$$\limsup_{n\to\infty} \frac{1}{n/2^r} \log P\{L_{n,r,k} \in F\} \le -I(F),$$

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

$$\liminf_{n\to\infty} \frac{1}{n/2^r} \log P\{L_{n,r,k} \in G\} \ge -I(G).$$

An example of such a sequence $L_{n,r,k}$ is given in Example 4.4. Finally, let $\mathscr{X}$ be a Polish space, θ a probability measure on $\mathscr{X}$ and $\Xi_r \doteq \{D_{r,k}, k = 1, \ldots, 2^r\}$ a partition of $\mathscr{X}$ satisfying Condition 2.1. The process whose asymptotics we will analyze in this section is the doubly indexed sequence

$$W_{n,r}(dx \times dy) \doteq \theta(dx) \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x) L_{n,r,k}(dy), \tag{4.1}$$

which takes values in $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$. As noted in Remark 2.2, the assumptions on $L_{n,r,k}$ can be considerably weakened.

The two-parameter large deviation theorem and Laplace principle for $W_{n,r}$ are stated next. The function J in (4.2) coincides with the function given in Definition 2.3.

THEOREM 4.1. *Let $W_{n,r}$ be defined by* (4.1). *We assume that for each n and r in $\mathbb{N}$ $L_{n,r,1}, L_{n,r,2}, \ldots, L_{n,r,2^r}$ are independent, that there exists a convex* *rate function I such that for each $k \in \{1, \dots, 2^r\}$ $\{L_{n,r,k}, n \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with scaling constants $n/2^r$ and rate function I, and that the partitions $\Xi_r = \{D_{r,k}\}$ satisfy Condition* 2.1. *Then the following conclusions hold*:

(a) *For $\mu = \theta \otimes \tau \in \mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$*

$$J(\mu) \doteq \int_{\mathscr{X}} I(\tau(x, \cdot))\, \theta(dx) \tag{4.2}$$

is a convex rate function.

(b) *The sequence $W_{n,r}$ satisfies the two-parameter LDP on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function J in the following sense. For any closed subset F of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$*

$$\limsup_{r\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\{W_{n,r} \in F\} \le -J(F),$$

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

$$\liminf_{r\to\infty} \liminf_{n\to\infty} \frac{1}{n} \log P\{W_{n,r} \in G\} \ge -J(G).$$

(c) *The sequence $W_{n,r}$ satisfies the two-parameter Laplace principle on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function J in the following sense. For any bounded continuous function h mapping $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ into $\mathbb{R}$*

$$\lim_{r\to\infty} \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \exp[nh(\mu)] P(W_{n,r} \in d\mu) = \sup_{\mu \in \mathscr{P}_\theta(\mathscr{X}\times\mathscr{Y})} \{h(\mu) - J(\mu)\}.$$

As in Corollary 2.6, Theorem 4.1 implies a number of related asymptotic results, which we will not state. The Laplace principle in part (c) of Theorem 4.1 is a direct consequence of the LDP in part (b) and is proved exactly as in the one-parameter case ([6], Section 1.2). Since the proofs of parts (a) and (b) of Theorem 4.1 are completely analogous to the proof of Theorem 2.4, we restrict our comments, omitting all details.

The proof of Theorem 4.1 relies on a number of lemmas. Of these, Lemmas 3.1, 3.2, 3.4, and 3.5 remain valid without change in the present setting. Lemma 3.3 is replaced by the following. It is a consequence of Lemmas 2.5, 2.7, 2.8 in [14] and the assumptions that for each n and r, $L_{n,r,1}, \dots, L_{n,r,2^r}$ are independent and that for each r and k, $\{L_{n,r,k}, n \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})$ with scaling constants $n/2^r$ and rate function I.

LEMMA 4.2. *For each r the sequence $\{(L_{n,r,1}, \dots, L_{n,r,2^r}), n \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Y})^{2^r}$ with scaling constants $n/2^r$ and rate function*

$$(\nu_1, \dots, \nu_{2^r}) \mapsto \sum_{k=1}^{2^r} I(\nu_k).$$

The proof of the large deviation lower bound in Theorem 4.1 is similar to the proof in Section 3.2; Lemma 4.2 is used in place of Lemma 3.3. Lemma 3.8, an exponential tightness result used to prove Theorem 2.4, is replaced by the following, the proof of which is omitted.

LEMMA 4.3. *Fix $r \in \mathbb{N}$. Then for each $M \in (0, \infty)$ there exists a compact subset K_M of $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ such that*

$$\limsup_{n\to\infty} \frac{1}{n/2^r} \log P\{W_{n,r} \in (K_M)^c\} \le -M.$$

The proof of the large deviation upper bound in Theorem 4.1 is similar to the proof in Section 3.4; Lemmas 4.2 and 4.3 are used in place of Lemmas 3.3 and 3.8. Finally, the proof that J has compact levels sets is similar to the proof in Section 3.5 and is omitted.

We end this section by giving an example of a sequence $L_{n,r,1}, \ldots, L_{n,r,2^r}$ satisfying the hypotheses of Theorem 4.1.

EXAMPLE 4.4. Each of the four cases considered in Example 2.7 gives rise to random measures $L_{n,r,1}, \ldots, L_{n,r,2^r}$ satisfying the hypotheses of Theorem 4.1. For simplicity we treat only the analogue of case (a). For each $r \in \mathbb{N}$ and $k \in \{1, \ldots, 2^r\}$, let $\{\alpha(n, r, k), n \in \mathbb{N}\}$ be a sequence satisfying

$$\lim_{n\to\infty} \frac{\alpha(n, r, k)}{n/2^r} = 1, \tag{4.3}$$

and for each n and r in $\mathbb{N}$ let $A(n, r, 1), \ldots, A(n, r, 2^r)$ be a sequence of disjoint subsets of $\mathbb{N}$ such that $\operatorname{card}(A_{n,r,k}) = \alpha(n, r, k)$. For example, define

$$A_{n,r,k} \doteq \left\{ i \in \mathbb{N}\colon \frac{n(k-1)}{2^r} < i \le \frac{nk}{2^r} \right\} \quad \text{and} \quad \alpha(n, r, k) \doteq \operatorname{card}(A_{n,r,k}).$$

In this case (4.3) is valid because

$$\frac{n}{2^r} - 1 \le \alpha(n, r, k) \le \frac{n}{2^r} + 1.$$

Now let $\mathscr{Z}$ be a Polish space and $\{\zeta_i, i \in \mathbb{N}\}$ a sequence of i.i.d. random variables taking values in $\mathscr{Z}$ and having common distribution ρ. For each n and r in $\mathbb{N}$ we define for $k \in \{1, \ldots, 2^r\}$ the empirical measures

$$L_{n,r,k} \doteq \frac{1}{\alpha(n, r, k)} \sum_{i \in A(n,r,k)} \delta_{\zeta_i},$$

which take values in $\mathscr{P}(\mathscr{Z})$. Then $L_{n,r,1}, \ldots, L_{n,r,2^r}$ are independent, and for each r and k Sanov's theorem implies that $\{L_{n,r,k}, n \in \mathbb{N}\}$ satisfies the LDP on $\mathscr{P}(\mathscr{Z})$ with scaling constants $\alpha(n, r, k)$ and the convex rate function $I \doteq R(\cdot \mid \rho)$. Since $\alpha(n, r, k)/(n/2^r) \to 1$ as $n \to \infty$, $\{L_{n,r,k}, n \in \mathbb{N}\}$ also satisfies the LDP on $\mathscr{P}(\mathscr{Z})$ with scaling constants $n/2^r$. According to Theorem 4.1, $W_{n,r}$ defined with these $L_{n,r,k}$ satisfies the two-parameter LDP and Laplace principle on $\mathscr{P}_\theta(\mathscr{X} \times \mathscr{Y})$ with rate function $J(\mu) = R(\mu \mid \theta \times \rho)$. This completes the example.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[2] BOUCHER, C., ELLIS, R. S. and TURKINGTON, B. (1998). Derivation of maximum entropy principles in two-dimensional turbulence via large deviations. Preprint.

[3] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Their Applications*. Jones and Bartlett, Boston.

[4] DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large Deviations*. Academic Press, Boston.

[5] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth and Brooks/Cole, Pacific Grove, CA.

[6] DUPUIS, P. and ELLIS, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley, New York.

[7] EISELE, T. and ELLIS, R. S. (1983). Symmetry breaking and random waves for magnetic systems on a circle. *Z. Wahrsch. Verw. Gebiete* **63** 279–348.

[8] ELLIS, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York.

[9] ELLIS, R. S. (1988). Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Ann. Probab.* **16** 1496–1508.

[10] ELLIS, R. S. (1995). An overview of the theory of large deviations and applications to statistical mechanics. *Scand. Actuar. J.* **1** 97–142.

[11] ELLIS, R. S. and WYNER, A. D. (1989). Uniform large deviation property of the empirical process of a Markov chain. *Ann. Probab.* **17** 1147–1151.

[12] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[13] FÖLLMER, H. and OREY, S. (1987). Large deviations for the empirical field of a Gibbs measure. *Ann. Probab.* **16** 961–977.

[14] LYNCH, J. and SETHURAMAN, J. (1987). Large deviations for processes with independent increments. *Ann. Probab.* **15** 610–627.

[15] MICHEL, J. and ROBERT, R. (1994). Large deviations for Young measures and statistical mechanics of infinite dimensional dynamical systems with conservation law. *Comm. Math. Phys.* **159** 195–215.

[16] MILLER, J. (1990). Statistical mechanics of Euler equations in two dimensions. *Phys. Rev. Lett.* **65** 2137–2140.

[17] OLLA, S. (1988). Large deviations for Gibbs random fields. *Probab. Theory Related Fields* **77** 343–359.

[18] ROBERT, R. (1989). Concentration et entropie pour les mesures d'Young." *C. R. Acad. Sci. Paris Sér. I* **309** 757–760.

[19] ROBERT, R. (1991). A maximum-entropy principle for two-dimensional perfect fluid dynamics. *J. Statist. Phys.* **65** 531–553.

[20] SHIRYAEV, A. N. (1996). *Probability*, 2nd ed. Springer, New York.

[21] STROOCK, D. W. (1984). *An Introduction to the Theory of Large Deviations*. Springer, New York.

[22] TURKINGTON, B. (1999). Statistical equilibrium measures and coherent states in two-dimensional turbulence. *Comm. Pure Appl. Math.* To appear.

C. BOUCHER
DEPARTMENT OF MATHEMATICS
ILLINOIS WESLEYAN UNIVERSITY
BLOOMINGTON, ILLINOIS 61702

R. S. ELLIS
B. TURKINGTON
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003
E-MAIL: rsellis@math.umass.edu