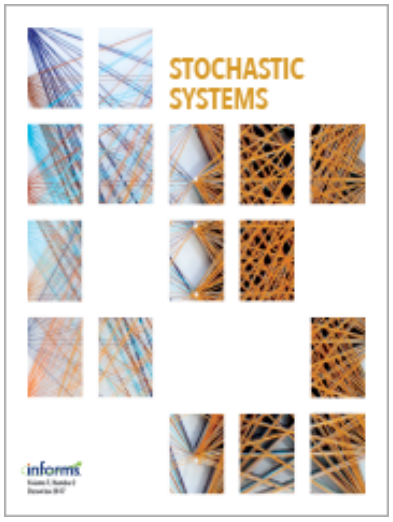
This article was downloaded by: [216.73.217.172] On: 15 November 2026, At: 06:26
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Asymptotics of the Invariant Measure in Mean Field Models with Jumps

Vivek S. Borkar, Rajesh Sundaresan

To cite this article:
Vivek S. Borkar, Rajesh Sundaresan (2012) Asymptotics of the Invariant Measure in Mean Field Models with Jumps. Stochastic Systems 2(2):322-380. https://doi.org/10.1287/12-SSY064

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "*Stochastic Systems*. Copyright 2012 The Author(s). https://doi.org/10.1287/12-SSY064, used under a Creative Commons Attribution License: http://creativecommons.org/licenses/by/4.0/."

Copyright © 2012, The author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Stochastic Systems
2012, Vol. 2, No. 2, 322–380
DOI: 10.1214/12-SSY064

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

ASYMPTOTICS OF THE INVARIANT MEASURE IN MEAN FIELD MODELS WITH JUMPS*

By Vivek S. Borkar† and Rajesh Sundaresan‡

Indian Institute of Technology Bombay and Indian Institute of Science

We consider the asymptotics of the invariant measure for the process of the empirical spatial distribution of N coupled Markov chains in the limit of a large number of chains. Each chain reflects the stochastic evolution of one particle. The chains are coupled through the dependence of the transition rates on this spatial distribution of particles in the various states. Our model is a caricature for medium access interactions in wireless local area networks. It is also applicable to the study of spread of epidemics in a network. The limiting process satisfies a deterministic ordinary differential equation called the McKean-Vlasov equation. When this differential equation has a unique globally asymptotically stable equilibrium, the spatial distribution asymptotically concentrates on this equilibrium. More generally, its limit points are supported on a subset of the ω-limit sets of the McKean-Vlasov equation. Using a control-theoretic approach, we examine the question of large deviations of the invariant measure from this limit.

1. Introduction. Spurred by the seminal work of Bianchi [5], there has been a flurry of activity in the communication networks community on mean field models for carrier sense multiple access (CSMA) protocols and their large time behavior. The continuous-time model for the wireless local area network (WLAN) is as follows. There are N particles (nodes) in the network. At each instant of time, a particle's state is a particular value taken from the finite state space $\mathcal{Z} = \{0, 1, \ldots, r-1\}$. A particle's

Received January 2012.

*A part of this paper without proofs was presented as an invited paper at the 2011 Annual Allerton Conference on Communication, Control, and Computing, Allerton, IL, USA, September 2011.

†Work supported by a J. C. Bose Fellowship and by IFCPAR (Indo-French Centre for the Promotion of Advanced Research), Project 4000-IT-1.

‡Work supported by IFCPAR (Indo-French Centre for the Promotion of Advanced Research), Project 4000-IT-1.

AMS 2000 subject classifications: Primary 60K35, 60F10, 68M20, 90B18, 49J15, 34H05.

Keywords and phrases: Decoupling approximation, fluid limit, invariant measure, McKean-Vlasov equation, mean field limit, small noise limit, stationary measure, stochastic Liouville equation.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

state represents the number of failed attempts at transmission of the head-of-the-line packet at that particle's queue. When a particle is in state i, a successful transmission gets the packet out of the system, and the particle moves to state 0 to service the next packet. A failed transmission moves the particle to state $i+1 \pmod r$. In the case when i was initially $r-1$, that is, $r-1$ unsuccessful transmission attempts were already made, another failed attempt results in the discarding of the packet. The particle then moves to state 0 with the next packet readied for transmission. We may interpret r as the maximum number of transmission attempts. The transition rate for a particle from state i to state j is governed by *mean field dynamics*, that is, the transition rate is $\lambda_{i,j}(\mu_N(t))$ where $\mu_N(t)$ is the empirical distribution of the states of particles at time t. If $X_n^{(N)}(t)$ is the state of the nth particle at time t, then one may write $\mu_N(t)$ as

$$\mu_N(t) = \frac{1}{N}\sum_{n=1}^{N} \delta_{\{X_n^{(N)}(t)\}}.$$

The particles interact only through the dependence of their transition rates on the current empirical measure $\mu_N(t)$.

The transitions allowed in the above model are from state i to either $i+1 \pmod r$ or 0. Let us say that $\mathcal{E}$ denotes the set of allowed transitions. In the above model,

$$\mathcal{E} = \{(i, i+1), i = 0, 1, \ldots, r-1\} \cup \{(i, 0), i = 0, 1, \ldots, r-1\}$$

where the addition is taken modulo r.

The process $X^{(N)}(\cdot) = \{X_n^{(N)}(\cdot), 1 \le n \le N\}$ is clearly a Markov process. But one difficulty needs to be surmounted in analyzing this system: the size of the state space grows exponentially in the number of particles. A step towards addressing this difficulty is to consider the evolution or flow of the empirical measure over time, which we shall call *empirical process*. This is a stochastic Liouville equation that lives on a smaller state space. In the infinite particle limit, this evolution turns out to be deterministic and is given by the McKean-Vlasov equation, whose large time behavior is an indicator of what one might expect of a finite but large population. In particular, if the deterministic evolution, given by the McKean-Vlasov equation, has a unique globally asymptotically stable equilibrium, then the states of a finite number of tagged particles are asymptotically independent and their joint law is given by the product of this equilibrium measure. The idea in fact was introduced by Kac as a simple model in kinetic theory [23] and was later studied by McKean and others (see, e.g., [25]). See [32] for an extensive account and [22] for a treatment of processes with jumps.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Several papers have provided rigorous analyses, along the above lines, of Bianchi's heuristic for studying WLANs. See, e.g., [9, 27, 8, 33, 3, 26]. See [16] for an excellent survey, [31, 1, 2] for early precursors, and [4] for an application of the same technique in game theory. As remarked in [16], experimental evidence for CSMA protocols indicates that the model and the predictions made by the analyses are surprisingly accurate even for small populations. Indeed, this is one of the main reasons for the model's enormous popularity. Is there a justification for this concentration phenomenon?

The mean-field analysis has been successful when the McKean-Vlasov equation has a unique globally asymptotically stable equilibrium. This is indeed the case in the simplest of WLAN settings with exponential backoff parameters. But there are settings with multiple stable equilibria [33] or with a unique equilibrium that is not globally stable (see [3] for a malware propagation model). In both cases, the dynamics governed by the McKean-Vlasov equation has multiple ω-limit sets. Significant effort has gone into identifying sufficient conditions for a unique equilibrium [26], and into identifying further sufficient conditions for a unique globally asymptotically stable equilibrium [8]. If there are multiple ω-limit sets for the McKean-Vlasov dynamics, which of these characterize the limiting behavior?

As a step in the direction of understanding these questions, we study the continuous-time model in this paper with the following goals.

1. Obtain a large deviation principle over finite time durations for the sequence of empirical measures and empirical processes, uniformly in the initial condition. This is of course rough asymptotics for large N, but suggests exponentially fast convergence to the deterministic limit.
2. Obtain a large deviation principle for the sequence of invariant measures, with single or multiple stable limit sets. This helps resolve which of the several ω-limit sets, when there are several, will be selected in the large N limit. (Our work can be straightforwardly extended to study the nature of transitions and exit times from the neighborhood of one stable equilibrium to the neighborhood of another, and help understand metastable behavior in such systems. We do not pursue these here.)
3. Provide a control theoretic framework to solve the problem of invariant measure in order to expose its strength and limitation. As we will see, we can go quite some distance using this *deterministic* approach, but eventually need to study the noisy system for resolution of some degeneracies.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

It must be noted that the above continuous-time model does not perfectly capture all aspects in a WLAN. In particular, interactions and changes of states occur in discrete-time units of slots in WLANs, and multiple nodes may transit in one slot. Multiple transitions never occur, almost surely, in our continuous-time model. Nevertheless, if the discrete-time model's transition rates and the slot sizes are appropriately scaled down as N grows so that the transition rates approach constants, our continuous time model provides accurate predictions of behavior on the discrete-time model. Our model also has wider applicability, one example being the study of spread of epidemics in a network; see [15, Sec. 2.4].

Large deviation principles over compact time durations for interacting diffusions and interacting jump Markov processes have been well-studied by several authors, e.g., [11, 24, 15, 19, 10, 14]. The works [15, 10, 14] establish large deviation principles for empirical measures in path space over finite time durations and characterize the rate functions, while [19] considers infinite time durations under the inductive topology. The works [11, 24, 15, 10], and [20] also study large deviation of the empirical process from the McKean-Vlasov limit, over finite time durations or infinite time durations under the inductive topology ([20]). The rate function measures the difficulty of passage of the empirical process in the neighborhood of a deviating path.

For a fixed N, when $t \to +\infty$, the stationary or invariant measure is of interest. When the limiting McKean-Vlasov dynamics has a unique globally asymptotically stable equilibrium, the invariant measure converges weakly in the infinite particle limit to the point mass at the equilibrium of the McKean-Vlasov dynamics (see, e.g., [3]). When there are multiple equilibria, the invariant measure concentrates on a subset of the ω-limit sets for the McKean-Vlasov dynamics ([21, Ch. 6] and [4]). Large deviations from this limit have been well-studied (see [21, Ch. 6], [30]). As indicated earlier, large deviation results for the invariant measure help resolve which of several possible ω-limit sets may be selected in the limit of a large number of nodes. They also help understand and predict metastable behavior in such systems. If the system is trapped in an undesirable equilibrium, exit times from domains and likely paths can be predicted.

Our approach for solving the large deviations of the invariant measure exploits a control-theoretic view described in [6], which considers a class of diffusions and generalizes results of [29] and [12] on small noise asymptotics of invariant measures and exit probabilities in diffusions. See [28] for small noise asymptotics of exit probabilities in processes with jumps. The control-theoretic approach differs from those of Freidlin and Wentzell [21]

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

and Shwartz and Weiss [30]. The latter rely on a study of an embedded Markov chain of states at hitting times of neighborhood of the stable limit sets (see, e.g., [21, Ch. 6.4]). Our control-theoretic approach enables the identification of a unique rate function when there is a unique globally asymptotically stable equilibrium for the McKean-Vlasov dynamics. When dealing with multiple ω-limit sets though, our approach eventually requires the study of the noisy system associated with the embedded Markov chain described above, for resolution of certain boundary conditions needed for a full characterization of the rate function.

We now outline our main arguments and describe the paper's organization.

We begin with a formal description of the model and statements of the main results on invariant measures in Section 2. In Section 3.1, we first establish a large deviation principle for empirical measures of paths over finite durations. For pure jump processes with interactions, this was established by [24, 15, 19, 10, 14], as indicated earlier. While [24] considers a fixed initial condition for all the particles, [10, 19], and [14] consider random, independent, and identically distributed (also called chaotic) initial conditions for the particles. To establish the large deviation principle for the interacting system, they first establish the large deviation principle for the independent and noninteracting system via Sanov's theorem, then exploit the Girsanov transformation to describe the probability measure for the interacting case, and then apply the Laplace-Varadhan principle. In order to eventually pass to the invariant measure, we need to establish a stronger uniform large deviation principle when the initial conditions of the particles are such that the initial empirical measures converge weakly, but are otherwise arbitrary. (See the remark following [19, Th. 4.1].) The limiting (initial) empirical measure defines the initial condition for the limiting deterministic McKean-Vlasov dynamics. This stronger uniform large deviation principle alluded to above is available for diffusions with mean field interactions in [11] and for certain classes of jump processes in [15] where the holding times alone are mediated by the interaction and not the jump probabilities. Our result is a mild extension facilitated by a generalization of Sanov's theorem given in [11]. While this part of our result is not surprising, we could not find a ready reference in the literature, and so we state the result in Section 3.1 and provide a proof in Section 5 based on the approach in [24]. We reemphasize that this result is only for finite durations, and is only a step towards addressing our next goal of asymptotics of the invariant measure.

In Section 3.2, we apply the contraction principle to obtain results on a large deviation from the McKean-Vlasov limit. As in [24], we prove this

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

under the finer uniform norm topology. In Section 3.3, we once again apply the contraction principle to argue a large deviation principle for terminal measure. We then establish some crucial estimates on the rate function for use in later sections. Finally in this section, we argue that if the initial measures satisfy a large deviation principle, then so do the joint initial and terminal measures.

In Section 4, we prove the large deviation principle for the invariant measure. To do this, we first establish a subsequential large deviation principle, then argue that the rate function satisfies the dynamic programming equation for a particular control problem, and then finally show that the rate function is unique if the associated McKean-Vlasov dynamics has a unique globally asymptotically stable equilibrium. The arguments to establish the uniqueness of the rate function parallel those of [6] with the significant difference that while the dynamic programming equation was arrived at in [6] via the theory of viscosity solutions, here we use the contraction principle. While this provides a complete result in case of a single globally asymptotically stable equilibrium, it leaves some indeterminacy regarding uniqueness of the so called 'potential function' whose global minima the invariant measure concentrates on in the large N limit. To resolve this, one has to fall back upon the framework of Freidlin and Wentzell [21, Ch.6] with minor modifications. These modifications are detailed in the Appendix.

Proofs that are not central to our control-theoretic view point are relegated to Sections 5, 6, 7, and 8. The Appendix details the minor modifications to the arguments of Freidlin and Wentzell [21, Ch.6] for resolution of the rate function values at the stable limit points.

2. The model and main results. Consider N interacting Markov chains denoted by

$$X_n^{(N)}(t),\ 1 \le n \le N,\ t \ge 0,$$

on a finite state space $\mathcal{Z} = \{0, \dots, r-1\}$, with dynamics as follows. $X_n^{(N)}(t)$ denotes the state of the nth particle at time t. Let $\mu_N(t)$ be the empirical measure of the particles at time t, that is,

$$\mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{\{X_n^{(N)}(t)\}} \in \mathcal{M}_1(\mathcal{Z}),$$

where $\mathcal{M}_1(\mathcal{Z})$ is the set of probability vectors over $\mathcal{Z}$, endowed with the topology of weak convergence on $\mathcal{M}_1(\mathcal{Z})$. A particle n in state i transits to state $j \in \mathcal{Z}$ with a rate $\lambda_{i,j}(\mu_N(t))$ that depends on the states of the other particles only through the empirical measure at that time. Thus the

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

processes interact only through the dependence of their transition rates on the current empirical measure.

Let $\mathcal{E} \subset \mathcal{Z} \times \mathcal{Z} \setminus \{(i,i) \mid i \in \mathcal{Z}\}$ be the set of admissible jumps for each particle. Thus $\lambda_{i,j}(\cdot) \equiv 0$ whenever $(i,j) \notin \mathcal{E}$ and $i \neq j$. We write $\mathcal{Z}_i = \{j \in \mathcal{Z} \mid (i,j) \in \mathcal{E}\}$ for the set of states to which a particle can jump from state i.

We make the following assumptions throughout the paper.

(**A1**) The graph with vertices $\mathcal{Z}$ and directed edges $\mathcal{E}$ is irreducible.

(**A2**) The mappings $\mu \in \mathcal{M}_1(\mathcal{Z}) \mapsto \lambda_{i,j}(\mu) \in [0,+\infty)$ are Lipschitz.

(**A3**) The rates of the admissible jumps are uniformly bounded away from zero, that is, there exists $c > 0$ such that, for all $\mu \in \mathcal{M}_1(\mathcal{Z})$ and all $(i,j) \in \mathcal{E}$, we have $\lambda_{i,j}(\mu) \geq c$.

Since $\mathcal{M}_1(\mathcal{Z})$ is compact and $\lambda_{i,j}(\cdot)$ are continuous by assumption (**A2**), the rates are uniformly bounded from above, that is, there is a $C < \infty$ such that for all $\mu \in \mathcal{M}_1(\mathcal{Z})$, and all $(i,j) \in \mathcal{E}$, we have $\lambda_{i,j}(\mu) \leq C$.

For any $T \in (0,+\infty)$, write $X_n^{(N)} : [0,T] \to \mathcal{Z}$ for the process of evolution of particle n over time. This is an element of the set $D([0,T],\mathcal{Z})$ of all cadlag paths from $[0,T]$ to $\mathcal{Z}$ equipped with the Skorohod topology. We set the path to be left continuous at T. Let

$$X^N = (X_n^{(N)}, 1 \leq n \leq N) \in D([0,T], \mathcal{Z}^N)$$

denote the full description of paths of all N particles. The initial condition at time $t = 0$ is $z^N = (z_n, 1 \leq n \leq N)$. The process X^N, with its law denoted $\mathbb{P}_{z^N}^{(N)}$, is a Markov process with cadlag paths, state space $\mathcal{Z}^N$, and generator $\mathscr{A}^{(N)}$ acting on bounded measurable functions Φ according to

$$\mathscr{A}^{(N)}\Phi(a^N) = \sum_{n=1}^{N} \sum_{j \in \mathcal{Z}_{a_n}} \left[\lambda_{a_n,j}\left(g_N(a^N)\right)\right] \left(\Phi(\sharp(a^N,n,j)) - \Phi(a^N)\right) \tag{2.1}$$

where $a^N = (a_n, 1 \leq n \leq N) \in \mathcal{Z}^N$, $\sharp(a^N,n,j)$ is the element of $\mathcal{Z}^N$ that results from replacing the nth component of a^N with j, and $g_N(a^N) = \frac{1}{N}\sum_{n'=1}^{N} \delta_{a_{n'}}$ is the empirical measure associated with the configuration a^N. (The upper boundedness of the rates implies that the martingale problem associated with the generator $\mathscr{A}^{(N)}$, operating on bounded measurable functions, and the initial condition $z^N \in \mathcal{Z}^N$ admits a unique solution $\mathbb{P}_{z^N}^{(N)} \in \mathcal{M}_1(D([0,T],\mathcal{Z}^N))$; see for e.g., [17, Problem 4.11.15, p. 263]).

The empirical measure process associated with X^N is given by

$$\mu_N : t \in [0,T] \mapsto \mu_N(t) = \frac{1}{N}\sum_{n=1}^{N} \delta_{\{X_n^{(N)}(t)\}} \in \mathcal{M}_1^{(N)}(\mathcal{Z}),$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

where $\mathcal{M}_1^{(N)}(\mathcal{Z}) = \{g_N(a^N) \mid a^N \in \mathcal{Z}^N\} \subset \mathcal{M}_1(\mathcal{Z})$. As a consequence of (2.1), the process μ_N is itself a Markov process with cadlag paths, finite state space $\mathcal{M}_1^{(N)}(\mathcal{Z})$, and generator $\mathcal{A}^{(N)}$ acting on bounded measurable functions Ψ according to

(2.2)
$$\mathcal{A}^{(N)}\Psi(\xi) = N \sum_{i\in\mathcal{Z}} \sum_{j\in\mathcal{Z}_i} [\xi(i)\lambda_{i,j}(\xi)] \left(\Psi(\xi - N^{-1}\delta_i + N^{-1}\delta_j) - \Psi(\xi)\right).$$

Write $\lambda_{i,i}(\xi) = -\sum_{j'\neq i} \lambda_{i,j'}(\xi)$ and define

$$A_\xi = (\lambda_{i,j}(\xi))_{(i,j)\in\mathcal{Z}\times\mathcal{Z}} \tag{2.3}$$

to be the *rate matrix* over $\mathcal{Z}$. It is well known that the family $(\mu_N, N \geq 1)$ satisfies the weak law of large numbers in the following sense: if $\mu_N(0) \to \nu$ weakly as $N \to \infty$ for some $\nu \in \mathcal{M}_1(\mathcal{Z})$, then $\mu_N \to \mu$ uniformly on compacts in probability, where μ solves the *McKean-Vlasov equation*

$$\dot{\mu}(t) = A^*_{\mu(t)}\mu(t) \tag{2.4}$$

with initial condition $\mu(0) = \nu$. Here $\mu(t)$ is interpreted as a column vector and $A^*_{\mu(t)}$ is the adjoint/transpose of the matrix $A_{\mu(t)}$. By assumption (**A2**), uniqueness of solutions holds for the nonlinear ordinary differential equation (ODE) (2.4).

As a consequence of the assumption (**A1**), for each $N \geq 1$, there is a unique invariant measure for the Markov process X^N, and hence there is a unique invariant measure $\wp^{(N)}$ for the $\mathcal{M}_1^{(N)}(\mathcal{Z})$-valued Markov process μ_N.

Define $\tau : \mathbb{R} \to \mathbb{R}_+$ to be

$$\tau(u) = e^u - u - 1, \tag{2.5}$$

and let $\tau^* : \mathbb{R} \to \overline{\mathbb{R}}_+$ be its Legendre conjugate

$$\tau^*(u) = \begin{cases} (u+1)\log(u+1) - u & \text{if } u > -1 \\ 1 & \text{if } u = -1 \\ +\infty & \text{if } u < -1. \end{cases} \tag{2.6}$$

The first main result is on the large N asymptotics of the sequence of invariant measures $(\wp^{(N)}, N \geq 1)$.

THEOREM 2.1. *Assume* (**A1**) - (**A3**). *Let the McKean-Vlasov equation* $\dot{\mu}(t) = A^*_{\mu(t)}\mu(t)$ *have a unique globally asymptotically stable equilibrium* ξ_0.

Then the sequence $(\wp^{(N)}, N \geq 1)$ *satisfies the large deviation principle with speed* N *and good rate function* s *given by*

$$s(\xi) = \inf_{\hat{\mu}} \int_{[0,+\infty)} \Big[\sum_{(i,j)\in\mathcal{E}} (\hat{\mu}(t)(i))\lambda_{i,j}(\hat{\mu}(t))\tau^* \left(\frac{\hat{l}_{i,j}(t)}{\lambda_{i,j}(\hat{\mu}(t))} - 1 \right) \Big] \, dt \tag{2.7}$$

where the infimum is over all $\hat{\mu}$ *that are solutions to the dynamical system* $\dot{\hat{\mu}}(t) = -\hat{L}(t)^*\hat{\mu}(t)$ *for some family of rate matrices* $\hat{L}(\cdot)$*, with initial condition* $\hat{\mu}(0) = \xi$*, terminal condition* $\lim_{t\to+\infty} \hat{\mu}(t) = \xi_0$*, and* $\hat{\mu}(t) \in \mathcal{M}_1(\mathcal{Z})$ *for all* $t \geq 0$*.*

REMARKS. 1. As we shall see later, the dynamics $\dot{\hat{\mu}}(t) = -\hat{L}(t)^*\hat{\mu}(t)$ corresponds to a time reversal when compared with the direction of the McKean-Vlasov dynamics. The rate function is given by the cost, associated with a certain control problem, of the cheapest path (across all time) that transports the system state from the globally asymptotically stable equilibrium ξ_0 to ξ, in the forward-time dynamics proceeding in the direction of the McKean-Vlasov dynamics. In the time-reversed dynamics, this is the cost of a path with initial state ξ and terminal state ξ_0.

2. By setting $\hat{l}_{i,j}(t) \equiv \lambda_{i,j}(\xi_0)$, we see that if $\hat{\mu}(0) = \xi_0$, then the system state remains $\hat{\mu}(t) \equiv \xi_0$, and the integral is 0. It follows that $s(\xi_0) = 0$.

3. In reversed time, the system state under the dynamics $\dot{\hat{\mu}}(t) = -\hat{L}(t)^*\hat{\mu}(t)$ may not in general lie in $\mathcal{M}_1(\mathcal{Z})$. The minimization however is over all paths that are constrained to lie in $\mathcal{M}_1(\mathcal{Z})$. See additional remarks after Lemma 4.2.

4. We now make some remarks on why Theorem 2.1 and the soon to follow Theorem 2.2 are not subsumed by the works of [30] and [21]. The transition rates for the Markov process $\mu_N(\cdot)$ are $(\mu_N(t)(i))\lambda_{i,j}(\mu_N(t))$. While $\lambda_{i,j}(\mu_N(t))$ is indeed bounded away from zero when $(i,j) \in \mathcal{E}$, $(\mu_N(t)(i))\lambda_{i,j}(\mu_N(t))$ is not, and so the logarithm of these rates is not bounded, a requirement in [30]. Next, the process μ_N takes values only in $\mathcal{M}_1(\mathcal{Z})$, and so when at the boundary $\mathcal{M}_1(\mathcal{Z})$, it is constrained to move only in those directions that keep it within $\mathcal{M}_1(\mathcal{Z})$. As a consequence, a finiteness condition [21, p.146, II] on the associated Lagrangian function does not hold.

5. Theorem 2.1 provides a complete characterization of the rate function. However, numerical computation of the rate function is a challenging problem. One might possibly discretize time and the state space, and employ dynamic programming techniques to get an approximation. This is an interesting line of work that is beyond the scope of this paper. See [4] and references therein for some results based on exit times.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

We next state a generalization of Theorem 2.1 when there may be multiple ω-limit sets. Consider a time-varying rate matrix $L(t)$ associated with the time-varying rates $(l_{i,j}(t), (i,j) \in \mathcal{E})$ and let $L(t)^*$ be its adjoint. Write μ for the solution to the dynamical system $\dot{\mu}(t) = L(t)^*\mu(t)$ with initial condition $\mu(0) = \nu$. Define

$$S_{[0,T]}(\mu|\nu) = \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\lambda_{i,j}(\mu(t))\tau^* \left(\frac{l_{i,j}(t)}{\lambda_{i,j}(\mu(t))} - 1 \right) \Big] \, dt. \tag{2.8}$$

As we shall see later, this is the cost of moving the system state along the trajectory μ with initial state $\mu(0) = \nu$. Let $V(\xi|\nu)$ be the so-called quasipotential defined by

$$V(\xi|\nu) = \inf\{S_{[0,T]}(\mu|\nu) \mid \mu(0) = \nu, \mu(T) = \xi, t \in [0,T], T \geq 0\}, \tag{2.9}$$

that is, the infimum cost of traversal from ν to ξ over all finite time durations. Let us define an equivalence relation on $\mathcal{M}_1(\mathcal{Z})$ as follows. We say $\nu \sim \xi$ if $V(\xi|\nu) = V(\nu|\xi) = 0$. Using a later result (the third part of Lemma 3.2), it is easy to show that the set of points that are equivalent to each other is closed and therefore compact (being a closed subset of the compact set $\mathcal{M}_1(\mathcal{Z})$).

We will generalize Theorem 2.1 under the following assumption on the dynamical system corresponding to the McKean-Vlasov equation (2.4):

(**B**) There exist a finite number of compact sets $K_1, K_2, \ldots, K_l$ such that

1. $\nu_1, \nu_2 \in K_i$ implies $\nu_1 \sim \nu_2$.
2. $\nu_1 \in K_i, \nu_2 \notin K_i$ implies $\nu_1 \nsim \nu_2$.
3. Every ω-limit set of the McKean-Vlasov equation $\dot{\mu}(t) = A^*_{\mu(t)}\mu(t)$ is contained in one of the K_i.

Under the hypothesis of Theorem 2.1, assumption (**B**) holds with $l = 1$, $K_1 = \{\xi_0\}$. We saw in a remark following Theorem 2.1 that $s(\xi_0) = 0$. In the general case, we shall see later in Lemma 4.6 that the rate function s is constant within each K_i, and so let s_i be the value of $s(\cdot)$ over K_i, for $i = 1, \ldots, l$. In order to specify the values of $s_1, \ldots, s_l$, define

$$\tilde{V}(K_i, K_j) = \inf_{T>0} \big\{ S_{[0,T]}(\mu|\mu(0)) \mid \mu(0) \in K_i, \mu(T) \in K_j, \\ \mu(t) \notin \cup_{i' \neq i,j} K_{i'} \text{ for } t \in [0,T] \big\}. \tag{2.10}$$

If the set is empty, the infimum is taken to be $+\infty$. Let us also define

$$V(K_i, K_j) \quad = \quad V(\xi|\nu)|_{\nu\in K_i, \xi \in K_j}. \tag{2.11}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Again using Lemma 4.6, one can show that the above value is independent of $\nu \in K_i$ and $\xi \in K_j$. Also, as indicated by Freidlin and Wentzell ([21, p.171]) and by Lemma A.2 of the Appendix, one can easily argue that

$$\begin{aligned} V(K_i, K_j) &= \tilde{V}(K_i, K_j) \wedge \min_{i_1}[\tilde{V}(K_i, K_{i_1}) + \tilde{V}(K_{i_1}, K_j)] \\ &\quad \wedge \min_{i_1, i_2}[\tilde{V}(K_i, K_{i_1}) + \tilde{V}(K_{i_1}, K_{i_2}) + \tilde{V}(K_{i_2}, K_j)] \\ &\quad \wedge \cdots \wedge \min_{i_1, \ldots, i_{l-2}} [\tilde{V}(K_i, K_{i_1}) + \cdots + \tilde{V}(K_{i_{l-2}}, K_j)]. \end{aligned}$$

Consider the indices $\{1, 2, \ldots, l\}$ for the compact sets $K_1, K_2, \ldots, K_l$. Let $\mathbb{G}\{i\}$ be the set of all *directed graphs* on the vertex set $\{1, 2, \ldots, l\}$ such that

- there is no outward edge from i;
- a vertex $j \neq i$ has exactly one outward edge;
- there are no closed cycles in the graph.

Define

$$W(K_{i'}) = \min_{\mathcal{G} \in \mathbb{G}\{i'\}} \sum_{(i,j) \in \mathcal{G}} V(K_i, K_j), \tag{2.12}$$

and finally

$$s_{i'} = W(K_{i'}) - \min_i W(K_i). \tag{2.13}$$

We now state the generalization of Theorem 2.1 under assumption (**B**).

THEOREM 2.2. *Assume* (**A1**) - (**A3**) *and* (**B**). *The sequence* $(\wp^{(N)}, N \geq 1)$ *satisfies the large deviation principle with speed* N *and good rate function* s *given by*

$$s(\xi) = \inf_{l'} \inf_{\hat{\mu}} \left[s_{l'} + \int_0^{+\infty} \Big[\sum_{(i,j) \in \mathcal{E}} (\hat{\mu}(t)(i)) \lambda_{i,j}(\hat{\mu}(t)) \tau^* \left(\frac{\hat{l}_{i,j}(t)}{\lambda_{i,j}(\hat{\mu}(t))} - 1 \right) \Big] \, dt \right] \tag{2.14}$$

where the second infimum is over all $\hat{\mu}$ *that are solutions to the dynamical system* $\dot{\hat{\mu}}(t) = -\hat{L}(t)^* \hat{\mu}(t)$ *for some family of rate matrices* $\hat{L}(\cdot)$, *with initial condition* $\hat{\mu}(0) = \xi$, *terminal condition* $\hat{\mu}(t) \to K_{l'}$ *as* $t \to +\infty$, *and* $\hat{\mu}(t) \in \mathcal{M}_1(\mathcal{Z})$ *for all* $t \geq 0$.

3. Large deviations over a finite time duration. In the previous section, we stated the main results of the paper on the sequence of invariant measures. We now begin our journey towards the proofs by studying large deviation principles over finite time durations. We first study the empirical measure of paths over finite durations, and then the empirical measure process.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

3.1. *Empirical measure.* Recall that $x^N \in D([0,T], \mathcal{Z}^N)$ denotes the full description of all the N particles. Let G_N denote the mapping that takes the full description x^N to the empirical measure

$$G_N : (x_n, 1 \le n \le N) \in D([0,T], \mathcal{Z}^N) \mapsto \frac{1}{N} \sum_{n=1}^{N} \delta_{x_n} \in \mathcal{M}_1(D([0,T], \mathcal{Z})).$$

Given the random variable X^N, the random empirical measure, denoted M_N, is thus

$$M_N = G_N(X^N) \in \mathcal{M}_1(D([0,T], \mathcal{Z})).$$

Clearly, the law of M_N depends on the initial condition z^N only through its empirical measure $\nu_N = (1/N) \sum_{n=1}^{N} \delta_{z_n}$. Write $P_{\nu_N}^{(N)}$ for the law of M_N, the push forward of $\mathbb{P}_{z^N}^{(N)}$ under the mapping G_N, that is, $P_{\nu_N}^{(N)} = \mathbb{P}_{z^N}^{(N)} \circ G_N^{-1}$.

The $\mathcal{M}_1^{(N)}(\mathcal{Z})$-valued cadlag empirical process is

$$\mu_N : t \in [0,T] \mapsto \mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{\{X_n^{(N)}(t)\}} \in \mathcal{M}_1^{(N)}(\mathcal{Z}),$$

and the corresponding mapping is denoted

$$\gamma_N : (x_n, 1 \le n \le N) \in D([0,T], \mathcal{Z}^N) \mapsto \mu_N : [0,T] \to \mathcal{M}_1^{(N)}(\mathcal{Z}).$$

Observe that $\mu_N(0) = \nu_N$, and that $\mu_N(t)$ is the projection $\pi_t(M_N)$ at time t, and we write $\mu_N = \pi(M_N)$. We thus have

$$\mu_N = \pi(M_N) = \pi(G_N(X^N)) = \gamma_N(X^N).$$

Consider now a hypothetical tagged particle. When there is no interaction, when all transition rates for $(i,j) \in \mathcal{E}$ transitions are unity, and when all other transition rates are 0, we can define the evolution of the tagged particle by the law P_z which is the unique solution to the martingale problem in $D([0,T], \mathcal{Z})$ associated with the generator A^o operating on bounded measurable functions Φ on $\mathcal{Z}$ according to

$$A^o \Phi(i) = \sum_{j \in \mathcal{Z}_i} 1 \cdot (\Phi(j) - \Phi(i))$$

and the initial condition z. The existence of a solution and the solution's uniqueness hold because the transition rates are upper bounded (see [17, Problem 4.11.15]). For any fixed $\mu \in D([0,T], \mathcal{M}_1(\mathcal{Z}))$, let $P_z(\mu)$ be the

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

unique solution to the martingale problem in $D([0,T],\mathcal{Z})$ associated with the (time-varying) generator

$$A_{\mu(t)}\Phi(i) = \sum_{j\in\mathcal{Z}_i} \lambda_{i,j}(\mu(t))\cdot(\Phi(j)-\Phi(i)) \tag{3.1}$$

and the initial condition z. This is consistent with (2.3) because when we view Φ and $A_\xi\Phi$ as column vectors, then the vector $A_\xi\Phi$ is the result of the rate matrix $A_\xi = (\lambda_{i,j}(\xi))_{(i,j\in\mathcal{Z}\times\mathcal{Z})}$ right-multiplied by the vector Φ. Again, by the upper boundedness of $\lambda_{i,j}(\cdot)$ (from assumptions (**A2**-**A3**)), $P_z(\mu)$ is unique, and the density of $P_z(\mu)$ with respect to P_z can be written as (see [24, eqn. (2.4)])

$$\frac{dP_z(\mu)}{dP_z}(x) = \exp\{h_1(x;\mu)\} \tag{3.2}$$

where

$$h_1(x;\mu) = \sum_{0\le t\le T} \mathbf{1}_{\{x_t\neq x_{t-}\}} \log\lambda_{x_{t-},x_t}(\mu(t-)) - \int_{[0,T]} \sum_{j\in\mathcal{Z}_{x_t}} (\lambda_{x_t,j}(\mu(t))-1)\ dt. \tag{3.3}$$

Consider the product distribution $\mathbb{P}^{o,(N)}_{z^N} = \bigotimes_{n=1}^N P_{z_n}$ where the N particles evolve independently, with the nth particle's initial condition being z_n. Again with $\nu_N = (1/N)\sum_{n=1}^N \delta_{z_n}$, let $P^{o,(N)}_{\nu_N} = \mathbb{P}^{o,(N)}_{z^N}\circ G_N^{-1}$. A simple application of Girsanov's formula yields that (see [24, eqn. (2.8)])

$$\frac{dP^{(N)}_{\nu_N}}{dP^{o,(N)}_{\nu_N}}(Q) = \exp\{Nh(Q)\} \tag{3.4}$$

where h is related to $h_1(\cdot;\cdot)$ as follows: for a $Q\in\mathcal{M}_1(D([0,T],\mathcal{Z}))$,

$$h(Q) = \int_{D([0,T],\mathcal{Z})} h_1(x;\pi(Q))\ Q(dx). \tag{3.5}$$

Let us define the spaces and topologies of interest. Similar to [24], consider the Polish space $(\mathcal{X},d)$ where

$$\mathcal{X} = \Big\{x\in D([0,T],\mathcal{Z}) \mid \sum_{0<t\le T}\mathbf{1}_{\{x(t)\neq x(t-)\}} < +\infty, \text{ and for each } t\in(0,T] \text{ with } x(t)\neq x(t-),\ \text{we have } x(t)\in\mathcal{Z}_{x(t-)}\Big\}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

with metric

$$d(x,y) = d_{\mathsf{Sko}}(x,y) + |\varphi(x) - \varphi(y)|, \quad x, y \in \mathcal{X}$$

where d_{Sko} stands for the Skorohod (complete) metric, and

$$\varphi : x \in \mathcal{X} \mapsto \sum_{0<t\leq T} \mathbf{1}_{\{x(t)\neq x(t-)\}} \in \mathbb{R}$$

denotes the number of jumps. φ is nonnegative and continuous (see [24, p. 299]). For a function $f : \mathcal{X} \to \mathbb{R}$, define

$$||f||_\varphi = \sup_{x\in\mathcal{X}} \frac{|f(x)|}{1+\varphi(x)},$$

denote

$$C_\varphi(\mathcal{X}) = \{f \mid f : \mathcal{X} \to \mathbb{R} \text{ is continuous and } ||f||_\varphi < +\infty\}$$

and

$$\mathcal{M}_{1,\varphi}(\mathcal{X}) = \left\{ Q \in \mathcal{M}_1(\mathcal{X}) \mid \int_{\mathcal{X}} \varphi \, dQ < +\infty \right\}.$$

This is a subset of the algebraic dual $C_\varphi(\mathcal{X})^*$ of $C_\varphi(\mathcal{X})$. Endow the set $\mathcal{M}_{1,\varphi}(\mathcal{X})$ with the weak* topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$, the weakest topology under which $Q_N \to Q$ as $N \to +\infty$ if and only if

$$\int_{\mathcal{X}} f \, dQ_N \to \int_{\mathcal{X}} f \, dQ \quad \text{for each } f \in C_\varphi(\mathcal{X}).$$

This topology is obviously finer than the topology of weak convergence in $\mathcal{M}_1(\mathcal{X})$.

For a measure $\nu \in \mathcal{M}_1(\mathcal{Z})$, let P be the mixture given by

$$dP(x) = \sum_{z\in\mathcal{Z}} \nu(z) dP_z(x). \tag{3.6}$$

and let $P(\mu)$ be the mixture given by

$$dP(\mu)(x) = \sum_{z\in\mathcal{Z}} \nu(z) dP_z(\mu)(x). \tag{3.7}$$

Define the relative entropy $H : \mathcal{M}_{1,\varphi}(\mathcal{X}) \to [0, +\infty]$ of Q (with respect to P) as

$$H(Q|P) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{dQ}{dP}\right) \, dQ & \text{if } Q \ll P \\ +\infty & \text{otherwise.} \end{cases} \tag{3.8}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Also define the function $J : \mathcal{M}_{1,\varphi}(\mathcal{X}) \to [0, +\infty]$

$$J(Q) = \sup_{f \in C_\varphi(\mathcal{X})} \left[\int_{\mathcal{X}} f \; dQ - \sum_{z \in \mathcal{Z}} \nu(z) \log \int_{\mathcal{X}} e^f \; dP_z \right]. \tag{3.9}$$

We first state a simple extension of [24, Th. 2.1]. See remarks following the statement on the nature of the extension.

THEOREM 3.1. *Suppose that the initial conditions $\nu_N \to \nu$ weakly. The sequence $(P^{(N)}_{\nu_N}, N \geq 1)$ satisfies the large deviation principle in the space $\mathcal{M}_{1,\varphi}(\mathcal{X})$, endowed with the weak* topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$, with speed N and good rate function $I(Q) = J(Q) - h(Q)$. Furthermore, for each $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$, $I(Q)$ admits the representation*

$$I(Q) = \begin{cases} H(Q|P(\pi(Q))) & \text{if } Q \circ \pi_0^{-1} = \nu \\ +\infty & \text{otherwise.} \end{cases} \tag{3.10}$$

REMARKS. This is a mild generalization of [24, Th. 2.1]. First, as in [24], the statement is stronger than usual statements pertaining to the topology of weak convergence since the topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$ is finer. Second, while [24] studied $z_n = z_0$ for some fixed z_0 so that $\nu_N = \delta_{z_0}$, we need to consider more general starting points for each particle, with the only proviso that the initial empirical measures ν_N converge weakly to ν. This generalization also goes beyond the chaotic initial conditions considered in [10, 19], and [14]. Third, there is another difference with [24] in that not all transitions are allowed, but only those in $\mathcal{E}$. To get the same results, irreducibility of the graph with vertices $\mathcal{Z}$ and directed edges $\mathcal{E}$ suffices (assumption (**A1**)).

PROOF. This is a straightforward extension of the proof of [24, Th. 2.1]. The arguments needed for the extension are highlighted in Section 5 for the sake of completeness. □

3.2. *Empirical process.* Recall the mapping

$$\gamma_N : (x_n, 1 \leq n \leq N) \in D([0,T], \mathcal{Z}^N) \mapsto \mu_N : [0,T] \to \mathcal{M}_1(\mathcal{Z}).$$

The flow μ_N takes values in the space $D([0,T], \mathcal{M}_1(\mathcal{Z}))$. Equip this space with the metric

$$\rho_T(\xi, \xi') = \sup_{0 \leq t \leq T} \rho_0(\xi_t, \xi'_t), \quad \xi, \xi' \in D([0,T], \mathcal{M}_1(\mathcal{Z})), \tag{3.11}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

where ρ_0 is taken to be a metric on $\mathcal{M}_1(\mathcal{Z})$ that metrizes the topology of weak convergence on $\mathcal{M}_1(\mathcal{Z})$. The space $D([0,T],\mathcal{M}_1(\mathcal{Z}))$ with the metric ρ_T is not separable. We are now interested in the law $p^{(N)}_{\nu_N}$ of μ_N which is the push forward $p^{(N)}_{\nu_N} = \mathbb{P}^{(N)}_{z^N} \circ \gamma_N^{-1}$. Since $\mu_N = \pi(M_N)$, we can also write $p^{(N)}_{\nu_N}$ as the push forward $p^{(N)}_{\nu_N} = P^{(N)}_{\nu_N} \circ \pi^{-1}$.

LEMMA 3.1. *The mapping $\pi : \mathcal{M}_{1,\varphi}(\mathcal{X}) \to D([0,T],\mathcal{M}_1(\mathcal{Z}))$ is continuous at each $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$ where $J(Q) < +\infty$. In particular, for each $t \in [0,T]$, the projection $\pi_t : \mathcal{M}_{1,\varphi}(\mathcal{X}) \to \mathcal{M}_1(\mathcal{Z})$ is continuous at each $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$ where $J(Q) < +\infty$.*

PROOF. The proof is included immediately after Lemma 5.8 in Section 5.2. □

Recall the definition of τ and its Legendre conjugate τ^* in (2.5) and (2.6), respectively. For $\theta : \mathcal{Z} \to \mathbb{R}$, $\xi \in \mathcal{M}_1(\mathcal{Z})$, define

$$|||\theta|||_\xi = \sup_{\Phi:\mathcal{Z}\to\mathbb{R}} \left\{ \sum_{i\in\mathcal{Z}} \theta(i)\cdot\Phi(i) - \sum_{(i,j)\in\mathcal{E}} \tau(\Phi(j)-\Phi(i))\cdot\xi(i)\cdot\lambda_{i,j}(\xi) \right\}.$$

The rate function of interest will be the following. For a $\nu \in \mathcal{M}_1(Z)$ and $\mu \in D([0,T],\mathcal{M}_1(Z))$, define

(3.12)
$$S_{[0,T]}(\mu|\nu) = \begin{cases} \int_{[0,T]} |||\dot\mu(t) - A^*_{\mu(t)}\mu(t)|||_{\mu(t)}\, dt & \text{if } \mu(0)=\nu \text{ and } \mu\in\mathcal{A} \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathcal{A}$ is the set of absolutely continuous functions on $[0,T]$. The connection between (3.12) and (2.8) will become clear in the following assertion.

THEOREM 3.2. *(a) Suppose that the initial conditions $\nu_N \to \nu$ weakly. Then the sequence $(p^{(N)}_{\nu_N}, N \ge 1)$ satisfies the large deviation principle in the space $D([0,T],\mathcal{M}_1(\mathcal{Z}))$ (under the topology induced by the metric ρ_T, see (3.11)) with speed N and good rate function $S_{[0,T]}(\mu|\nu)$.*

(b) If the path $\mu \in D([0,T],\mathcal{M}_1(\mathcal{Z}))$ has $S_{[0,T]}(\mu|\nu) < +\infty$, then $\mu \in \mathcal{A}$ and there exist rates $(l_{i,j}(t), t\in[0,T],\ (i,j)\in\mathcal{E})$ such that

- *$\dot\mu(t) = L(t)^*\mu(t)$ where $L(t)$ is the rate matrix associated with the time-varying rates $(l_{i,j}(t),(i,j)\in\mathcal{E})$ and $L(t)^*$ is its adjoint;*
- *the good rate function $S_{[0,T]}(\mu|\nu)$ is given by the equation (2.8).*

(c) Suppose that the following hold: $\mu \in \mathcal{A}$, $\mu(0) = \nu$, there exist time-varying rates $(l_{i,j}(t), t \in [0,T],\ (i,j) \in \mathcal{E})$ such that the associated rate matrix $L(t)$ satisfies $\dot{\mu}(t) = L(t)^*\mu(t)$, and the right-hand side of (2.8) is finite. Then the good rate function $S_{[0,T]}(\mu|\nu)$ evaluated at μ is given by (2.8).

PROOF. With the generalization available in Theorem 3.1, the proof (of all three statements) is identical to the proof of [24, Th. 3.1]. See also [15, Th. 2] for statements on the nature of the rate function. □

REMARKS. 1. Yet again, this is a mild generalization of [24, Th. 3.1], and of the results in [10, 20], and [14] that assume chaotic initial conditions. We allow any arbitrary sequence of initial conditions ν_N so long as $\nu_N \to \nu$ weakly. See a consequence in Corollary 3.1 below.

2. Observe from (2.8) that if the rate function $S_{[0,T]}(\mu|\nu) = 0$ then μ must be the unique solution to the McKean-Vlasov equation $\dot{\mu}(t) = A^*_{\mu(t)}\mu(t)$ with initial condition $\mu(0) = \nu$. (That the solution is unique follows from Lipschitz assumption (**A2**) which implies the well-posedness of the ODE (2.4)). The claim $p^{(N)}_{\nu_N} \to \delta_{\mu(\cdot)}$ follows.

3. When a path μ is such that $S_{[0,T]}(\mu|\nu) < +\infty$, the first bullet in the second statement of Theorem 3.2 says that there is a control (tilt), given by the rate matrix $L(t)$, such that the normal limiting trajectory under this control is μ. The expression (3.12) and the alternate expression (2.8) reinforce the notion that $S_{[0,T]}(\mu|\nu)$ is the *cost* of transporting the system along the trajectory μ.

4. For any finite $B \geq 0$ and any $\nu \in \mathcal{M}_1(\mathcal{Z})$, the set $\{\mu \in \mathcal{A} \mid S_{[0,T]}(\mu|\nu) \leq B\}$ is compact since $S_{[0,T]}(\cdot|\nu)$ is a good rate function.

The following is a straightforward corollary which shows that $S_{[0,T]}(\mu|\nu)$ is a rate function for a large deviation principle that holds uniformly in the initial point.

COROLLARY 3.1. *For any compact set $K \subset \mathcal{M}_1(\mathcal{Z})$, any closed set $F \subset D([0,T], \mathcal{M}_1(\mathcal{Z}))$, and any open set $G \subset D([0,T], \mathcal{M}_1(\mathcal{Z}))$, we have*

$$\limsup_{N\to\infty} \frac{1}{N} \log \sup_{\nu\in K} p^{(N)}_\nu \{\mu_N \in F\} \leq -\inf_{\nu\in K, \mu\in F} S_{[0,T]}(\mu|\nu), \tag{3.13}$$

$$\liminf_{N\to\infty} \frac{1}{N} \log \inf_{\nu\in K} p^{(N)}_\nu \{\mu_N \in G\} \geq -\sup_{\nu\in K} \inf_{\mu\in G} S_{[0,T]}(\mu|\nu). \tag{3.14}$$

PROOF. This is immediate from [13, Cor. 5.6.15] in conjunction with Theorem 3.2. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

3.3. *Empirical measure at initial and terminal times.* We first study the behavior of the empirical measure at terminal time alone. For $T \geq 0$, recall that the law of the empirical process μ_N with initial condition $\mu_N(0) = \nu_N$ (arising from z^N) is $p^{(N)}_{\nu_N} = \mathbb{P}^{(N)}_{z^N} \circ \pi^{-1}$. The empirical measure at terminal time T is $\mu_N(T)$; its law is the push forward $p^{(N)}_{\nu_N,T} = \mathbb{P}^{(N)}_{z^N} \circ \pi_T^{-1}$. The following result is an easy consequence of the contraction principle [13, Th. 4.2.1].

THEOREM 3.3. *Suppose that the initial conditions $\nu_N \to \nu$ weakly. Then the family $(p^{(N)}_{\nu_N,T}, N \geq 1)$ satisfies the large deviation principle in $\mathcal{M}_1(\mathcal{Z})$ with speed N and good rate function*

$$S_T(\xi|\nu) = \inf\{S_{[0,T]}(\mu|\nu) \mid \mu(0) = \nu, \mu(T) = \xi, \mu \in \mathcal{A}\}. \tag{3.15}$$

Furthermore $S_T(\xi|\nu)$ is bounded for all $\nu, \xi \in \mathcal{M}_1(\mathcal{Z})$, and we may restrict attention in the infimum to $\mu \in \mathcal{A}$ that also satisfy $S_{[0,T]}(\mu|\nu) < +\infty$. Moreover, the infimum is attained, and there exist $\overline{\mu} \in \mathcal{A}$ with $\overline{\mu}(0) = \nu$, $\overline{\mu}(T) = \xi$, and rates $(l_{i,j}(t), t \in [0,T],\ (i,j) \in \mathcal{E})$ with associated rate matrix $L(t)$ such that $\dot{\overline{\mu}}(t) = L(t)^* \overline{\mu}(t)$, $S_T(\xi|\nu) = S_{[0,T]}(\overline{\mu}|\nu)$, and $S_{[0,T]}(\overline{\mu}|\nu)$ satisfies (2.8).

PROOF. Recall that the metric on $D([0,T], \mathcal{M}_1(\mathcal{Z}))$ is ρ_T given in (3.11). It follows that the mapping $\mu \in D([0,T], \mathcal{M}_1(\mathcal{Z})) \mapsto \mu(T) \in \mathcal{M}_1(\mathcal{Z})$ is continuous. The statement on the validity of the large deviation principle with good rate function (3.15) follows by the contraction principle [13, Th. 4.2.1]. In the second statement of the next lemma, we show that $S_T(\xi|\nu)$ is bounded for all $\nu, \xi \in \mathcal{M}_1(\mathcal{Z})$. The goodness of the rate function $S_{[0,T]}(\mu|\nu)$ implies that the infimum in (3.15) is attained. The rest follow from Theorem 3.2. □

Let us now prove the boundedness and related properties of $S_T(\xi|\nu)$.

LEMMA 3.2. *The following statements hold.*

- *There exists a constant $C_1(T) < +\infty$ such that for any $\xi, \nu \in \mathcal{M}_1(\mathcal{Z})$, there is a piecewise linear and continuous path μ having constant velocity in each linear segment and $S_{[0,T]}(\mu|\nu) \leq C_1(T)$.*
- *For any $\xi, \nu \in \mathcal{M}_1(\mathcal{Z})$, we have $S_T(\xi|\nu) \leq C_1(T)$.*
- *There exists a constant $C_2 < +\infty$ such that for every $\varepsilon > 0$, there is a $\delta \in (0, \varepsilon)$ such that $\rho_0(\nu, \xi) < \delta$ implies $S_\varepsilon(\xi|\nu) \leq C_2 \varepsilon$.*

PROOF. The main idea is to show that the difficulty of passage near the neighborhood of a constant velocity straight line path is bounded. See Section 6. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

We next have a useful uniform continuity result.

LEMMA 3.3. *The mapping* $(\nu, \xi) \mapsto S_T(\xi|\nu)$ *is uniformly continuous.*

PROOF. See Section 7. □

Thus far, the only condition we imposed on the initial conditions were that $\nu_N \to \nu$ weakly. Let $\wp_0^{(N)}$ denote the law of the initial empirical measure $\mu_N(0)$ and let $\wp_{0,T}^{(N)}$ denote the joint law of $(\mu_N(0), \mu_N(T))$. We now consider $(\wp_{0,T}^{(N)}, N \geq 1)$.

THEOREM 3.4. *Suppose that the sequence* $(\wp_0^{(N)}, N \geq 1)$ *satisfies the large deviation principle with speed* N *and good rate function* $s : \mathcal{M}_1(Z) \to [0, +\infty]$. *Then the sequence of joint laws* $(\wp_{0,T}^{(N)}, N \geq 1)$ *satisfies the large deviation principle with speed* N *and good rate function*

$$S_{0,T}(\nu, \xi) = s(\nu) + S_T(\xi|\nu).$$

PROOF. See Section 8. □

Before we close this section, we state a useful result on the uniform continuity of the quasipotential in (2.9). This is analogous to Lemma 3.3, except that the time durations are finite but otherwise arbitrary.

LEMMA 3.4. *The mapping* $(\nu, \xi) \mapsto V(\xi|\nu)$ *is uniformly continuous.*

PROOF. See the last part of Section 7. □

4. Invariant measure: A control theoretic approach. Recall that by assumption (**A1**), for each $N \geq 1$, the finite-state continuous-time Markov chain μ_N is irreducible, and hence has a unique invariant measure, which we denoted $\wp^{(N)}$. In this section, we establish the large deviation principle for $(\wp^{(N)}, N \geq 1)$ as stated in Theorems 2.1 and 2.2.

The outline of our control theoretic approach is the following.

- In Lemma 4.1, we first establish a subsequential large deviation principle. We shall also establish, via the contraction principle, that the rate function s satisfies a dynamic programming equation (see (4.1)). This equation naturally suggests a control problem with an associated running cost.
- There will be multiple solutions to (4.1). But the rate function that we are after will satisfy a further condition. In Lemma 4.2, we shall

show the existence of one single optimal path of infinite duration and shall extract a recursive equation for s from (4.1), a further condition that the rate function must satisfy. The heart of the control-theoretic approach lies in this step.

- We then argue in Lemma 4.3 that this optimal path must end up within a set that is positively invariant to the time-reversed McKean-Vlasov dynamics.
- The above steps fix the rate function at all points outside the ω-limit sets, assuming the values at the ω-limit sets.
- Subsection 4.1 then argues that in case of a unique globally asymptotically stable equilibrium, the rate function is zero at the equilibrium. Subsection 4.2 falls back on the approach of Freidlin and Wentzell [21, Ch. 6] under assumption (**B**) to fix the values at the ω-limit sets. This fixes the rate function uniquely for all subsequential large deviation principles, and the main results follow.

We begin by establishing subsequential large deviation principles and the dynamic programming equation.

LEMMA 4.1. *For any sequence of natural numbers going to $+\infty$, there exists a subsequence $(N_k,\ k \geq 1)$ such that $(\wp^{(N_k)},\ k \geq 1)$ satisfies the large deviation principle with speed N_k and a good rate function s that satisfies*

$$s(\xi) = \inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} [s(\nu) + S_T(\xi|\nu)] \quad \text{for every } T > 0. \tag{4.1}$$

Furthermore, $s \geq 0$ and there exists a $\nu^* \in \mathcal{M}_1(\mathcal{Z})$ such that $s(\nu^*) = 0$.

PROOF. Since the topology on $\mathcal{M}_1(\mathcal{Z})$ with metric ρ_0 has a countable base, and because $\mathcal{M}_1(\mathcal{Z})$ is compact, by [13, Lem. 4.1.23], there is a subsequence $N_k \to +\infty$ of the given sequence such that $(\wp^{(N_k)}, N_k \geq 1)$ satisfies the large deviation principle with speed N_k and a good rate function $s : \mathcal{M}_1(\mathcal{Z}) \to [0, +\infty]$.

We now verify (4.1). Fix an arbitrary $T > 0$. By Theorem 3.4, with $\wp_0^{(N)} = \wp^{(N)}$, the invariant measure, the sequence of joint laws $(\wp_{0,T}^{(N_k)},\ k \geq 1)$ satisfies the large deviation principle along the subsequence $(N_k,\ k \geq 1)$ with speed N_k and good rate function $S_{0,T}(\nu, \xi) = s(\nu) + S_T(\xi|\nu)$. By the contraction principle, the sequence of terminal laws $(\wp_T^{(N_k)},\ k \geq 1)$ satisfies the large deviation principle along the subsequence with the good rate function

$$\inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} S_{0,T}(\nu, \xi) = \inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} [s(\nu) + S_T(\xi|\nu)]. \tag{4.2}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

But $\wp^{(N)}$ is invariant to time shifts which yields $\wp_T^{(N)} = \wp_0^{(N)} = \wp^{(N)}$. The infimum on the left-hand side of (4.2) must therefore evaluate to $s(\xi)$, which yields (4.1).

Rate functions are nonnegative and have infimum value of 0, that is, $s \geq 0$ and $\inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} s(\nu) = 0$. Since s is a good rate function, the infimum 0 is attained at some point; call it ν^*. The proof is now complete. □

As indicated earlier, there are multiple solutions to (4.1). Indeed, $s(\cdot) \equiv 0$ is one of them. In order to identify a further condition that s must satisfy, we now identify a control problem associated to (4.1). To do this, we shall now consider paths that are of time-duration mT that end at ξ. Since the terminal condition is fixed, it would be convenient to fix the terminal time as 0 and look at negative times; in particular, paths in the time interval $[-mT, 0]$ for $m \geq 1$. But then, we may reverse time and consider the dynamical system

$$\dot{\hat{\mu}}(t) = -\hat{L}(t)^* \hat{\mu}(t), \quad \hat{\mu}(0) = \xi, \quad t \in [0, mT], \quad m \geq 1.$$

We use hats as in $\hat{\mu}, \hat{l}_{i,j}, \hat{\lambda}_{i,j}, \hat{L}$ to denote quantities where time flows in the opposite direction with reference to the direction under the McKean-Vlasov dynamics. In particular,

$$\begin{aligned} \hat{\mu}(t) &= \mu(mT - t) \\ \hat{l}_{i,j}(t) &= l_{i,j}(mT - t), \text{for all } i, j \in \mathcal{Z} \\ \hat{L}(t) &= L(mT - t). \end{aligned}$$

for $t \in [0, mT]$. Also, for uniformity in notation, let $\hat{\lambda}_{i,j}(\cdot) = \lambda_{i,j}(\cdot)$ for all (i, j) pairs. One then views the $\hat{L}(t)$ above as the control at time t when the state is $\hat{\mu}(t)$ with cost function at time t given by

$$\hat{r}(\hat{\mu}(t), \hat{L}(t)) = \sum_{(i,j) \in \mathcal{E}} (\hat{\mu}(t)(i)) \hat{\lambda}_{i,j}(\hat{\mu}(t)) \tau^* \left(\frac{\hat{l}_{i,j}(t)}{\hat{\lambda}_{i,j}(\hat{\mu}(t))} - 1 \right).$$

Observe that the cost function is zero when $\hat{l}_{i,j}(t) = \hat{\lambda}_{i,j}(\hat{\mu}(t))$ for almost every t in the time duration of interest. The total cost is

$$\int_{[0,mT]} \left[\sum_{(i,j) \in \mathcal{E}} (\hat{\mu}(t)(i)) \hat{\lambda}_{i,j}(\hat{\mu}(t)) \tau^* \left(\frac{\hat{l}_{i,j}(t)}{\hat{\lambda}_{i,j}(\hat{\mu}(t))} - 1 \right) \right] dt. \tag{4.3}$$

which is simply $S_{[0,mT]}(\mu | \mu(0))$ of (2.8) as can be verified by a change of variable in (4.3) that takes t to $mT - t$.

The following lemma establishes the existence of one optimal path $\hat{\mu}$ of infinite duration, starting at ξ.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA 4.2. *For each $\xi \in \mathcal{M}_1(\mathcal{Z})$, there exists a path $\hat{\mu} : [0, +\infty) \to \mathcal{M}_1(\mathcal{Z})$ and a family of rate matrices $(\hat{L}(t), t \in [0, +\infty))$ such that $\hat{\mu}(\cdot)$ satisfies the ODE*

$$\dot{\hat{\mu}}(t) = -\hat{L}(t)^* \hat{\mu}(t), \quad t \in [0, +\infty) \tag{4.4}$$

with initial condition $\hat{\mu}(0) = \xi$, and

$$s(\xi) = s(\hat{\mu}(mT)) + \int_{[0,mT]} \hat{r}(\hat{\mu}(t), \hat{L}(t))\, dt \quad \text{for all } m \geq 1. \tag{4.5}$$

REMARK. While the McKean-Vlasov dynamics or the more general $\dot{\mu}(t) = L(t)^* \mu(t)$ with $L(t)$ being a rate matrix ensures that $\mu(t)$ lies within $\mathcal{M}_1(\mathcal{Z})$, this is not the case for the dynamics given by (4.4). Indeed, at the boundary of $\mathcal{M}_1(\mathcal{Z})$, viewed as a subset of $\mathbb{R}^r$, the velocity for the dynamics in (4.4) points towards a direction of immediate exit from $\mathcal{M}_1(\mathcal{Z})$. The state space for the dynamics of (4.4) is therefore not restricted to $\mathcal{M}_1(\mathcal{Z})$. However, the lemma assures us that the selected path $\hat{\mu}$ stays within the compact subset $\mathcal{M}_1(\mathcal{Z})$ for all time.

PROOF. Our approach to prove this is the following. We shall define a topology on a suitable subspace of paths of infinite duration, and then show that we can restrict attention to a compact subset. We shall then argue that there exists a nested sequence of decreasing compact subsets, each of which is nonempty and all of whose elements satisfy the desired properties. The intersection will then be nonempty to yield the desired path.

Step 1: We shall now restrict attention to paths that lie inside $\mathcal{M}_1(\mathcal{Z})$. For a path $\hat{\mu} : [0, +\infty) \to \mathcal{M}_1(\mathcal{Z})$, define its restrictions to $[0, mT]$ by

$$\hat{\mu} \mapsto \psi_m \hat{\mu}(\cdot) = \hat{\mu}^{(m)}(\cdot) : [0, mT] \to \mathcal{M}_1(\mathcal{Z})$$

which is the restriction of the path $\hat{\mu}$ to $[0, mT]$. Consider the space of paths of infinite duration with metric

$$\rho_\infty(\hat{\mu}, \hat{\eta}) = \sum_{m=1}^{\infty} 2^{-m} \left(\rho_{mT}(\psi_m \hat{\mu}, \psi_m \hat{\eta}) \wedge 1\right).$$

Obviously, ψ_m is continuous for each m.

We shall also consider reversed restrictions (denoted without hats) defined by

$$\mu^{(m)}(t) = \hat{\mu}^{(m)}(mT - t) = \hat{\mu}(mT - t), \quad t \in [0, mT]. \tag{4.6}$$

Fix a $B \in [0, +\infty)$, and consider the set

$$\Gamma_\infty = \left\{ \hat{\mu}(\cdot) : [0, +\infty) \to \mathcal{M}_1(\mathcal{Z}) \mid \sup_{m \geq 1} S_{[0,mT]}(\mu^{(m)} | \mu^{(m)}(0)) \leq B \right\}. \tag{4.7}$$

Also, with $\eta(t) = \hat{\eta}(mT - t)$ for $t \in [0, mT]$, define

$$\Gamma_m = \left\{ \hat{\eta} : [0, mT] \to \mathcal{M}_1(\mathcal{Z}) \mid S_{[0,mT]}(\eta | \eta(0)) \leq B \right\}.$$

Γ_∞ is compact. To see this, take an arbitrary infinite sequence $(\hat{\mu}_n, n \geq 1) \subset \Gamma_\infty$. Since $S_{[0,mT]}$ is a good rate function, using Lemma 7.2 and Lemma 7.1, it is easy to see that each Γ_m is compact, and so one can find an infinite subset $\mathbb{V}_1 \subset \mathbb{N}$ such that $(\psi_1 \hat{\mu}_n, n \in \mathbb{V}_1) \subset \Gamma_1$ converges. Take a further subsequence represented by the infinite subset $\mathbb{V}_2 \subset \mathbb{V}_1$ such that $(\psi_2 \hat{\mu}_n, n \in \mathbb{V}_2) \subset \Gamma_2$ converges. Continue this procedure and take the subsequence along the diagonal. This subsequence converges for every interval $[0, mT]$. For each t define $\hat{\mu}(t)$ to be the point-wise limit. Since for each m, we have $\psi_m \hat{\mu} \in \Gamma_m$, it follows that with $\mu^{(m)}$ defined as in (4.6), $\sup_{m \geq 1} S_{[0,mT]}(\mu^{(m)} | \mu^{(m)}(0)) \leq B$, and so $\hat{\mu} \in \Gamma_\infty$. Thus Γ_∞ is sequentially compact, and by virtue of its being a subset of a metric space, Γ_∞ is compact.

Fix $\xi \in \mathcal{M}_1(\mathcal{Z})$. Consider the time duration $[0, T]$. We know from Lemma 4.1 that there is a ν^* with $s(\nu^*) = 0$. Using this in (4.1) of Lemma 4.1, we get

$$s(\xi) \leq S_T(\xi | \nu^*) \leq C_1(T)$$

where the last inequality is due to the second statement in Lemma 3.2. Take the constant $B = C_1(T)$; the corresponding Γ_∞ is compact.

Step 2: Observe that (4.1) can be viewed as a minimization over path space, with paths μ of duration $[0, mT]$ ending at ξ. Starting from any initial location ν, the minimum value is upper bounded by B. Indeed, traverse the McKean-Vlasov path with initial condition ν for duration $(m-1)T$. This contributes zero to the cost. Then proceed to ξ in T units of time. This costs at most $B = C_1(T)$. The minimum cost to go from ν to ξ in time $[0, mT]$ is thus at most B. If we consider reversed and translated time so that initial time is 0, the reversed paths $\hat{\mu}$ begin at ξ at time 0, have cost at most B, and stay in $\mathcal{M}_1(\mathcal{Z})$ for the duration $[0, mT]$.

Let

$$\Gamma_m^* = \bigcup_{\nu \in \mathcal{M}_1(\mathcal{Z})} \Big\{ \hat{\mu} \mid \mu(t) = \hat{\mu}(mT - t), \mu(0) = \nu, \mu(mT) = \xi, \\ S_{[0,mT]}(\mu | \nu) = S_{mT}(\xi | \nu) \Big\},$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

that is, the collection of all minimum cost paths $\hat{\mu}$ in $[0, mT]$ from ξ to every location in $\mathcal{M}_1(\mathcal{Z})$; the minimum cost is at most B. Clearly, $\Gamma_m^* \subset \Gamma_m$, and $\Gamma_m^* \neq \emptyset, m \geq 1$.

Γ_m^* is also compact. This set being a subset of the compact set Γ_m, it suffices to show that Γ_m^* is closed. Let $\hat{\mu}$ be a point of closure of Γ_m^*. We can then find a sequence $(\hat{\mu}^{(k)}, k \geq 1) \subset \Gamma_m^*$ such that $\lim_{k\to+\infty} \hat{\mu}^{(k)} = \hat{\mu}$. Clearly, we must have $\hat{\mu}(0) = \xi$. Let $\nu = \hat{\mu}(mT)$. By a simple application of lower semicontinuity of $S_{[0,mT]}(\cdot|\nu)$, Lemma 7.2, and Lemma 7.1, we must have

$$\begin{aligned} S_{[0,mT]}(\mu|\nu) &\leq \liminf_{k\to+\infty} S_{[0,mT]}(\mu^{(k)}|\mu^{(k)}(0)) \\ &= \liminf_{k\to+\infty} S_{mT}(\xi|\mu^{(k)}(0)) \\ &= S_{mT}(\xi|\nu) \end{aligned}$$

where the last inequality follows because of the continuity of S_{mT} in its arguments. But $S_{mT}(\xi|\nu)$ is the least cost for paths that traverse from ν to ξ in duration $[0, mT]$. So we must have $S_{[0,mT]}(\mu|\nu) = S_{mT}(\xi|\nu)$, which establishes that $\mu \in \Gamma_m^*$; Γ_m^* is therefore closed.

Step 3: Let us now finish the proof of Lemma 4.2. Since Γ_m^* is nonempty and compact, the continuity of ψ_m implies that $\psi_m^{-1}\Gamma_m^*$ is nonempty and closed. Further, being a closed subset of the compact set Γ_∞, $\psi_m^{-1}\Gamma_m^*$ is itself compact. A simple dynamic programming argument further shows that $\psi_m^{-1}\Gamma_m^*$ is a nested decreasing sequence of subsets. Their intersection is nonempty. Take a $\hat{\mu}$ in the intersection.

Focusing on the duration $[mT, mT + T]$, the path $t \in [0, T] \mapsto \eta^{(m)}(t) = \hat{\mu}(mT + T - t)$ has $S_{[0,T]}(\eta^{(m)}|\eta^{(m)}(0)) \leq B < +\infty$, and so, by the last part of Theorem 3.2, we can find rates $L^{(m)}(t)$ such that $\eta^{(m)}$ satisfies $\dot{\eta}^{(m)}(t) = (L^{(m)}(t))^*\eta^{(m)}(t)$, with initial condition $\eta^{(m)}(0)$. Put these pieces of duration T together by defining $\hat{L}(mT + t) = L^{(m)}(T - t)$ for $t \in [0, T]$, and we get a rate matrix $\hat{L}(\cdot)$ defined on $[0, +\infty)$ such that $\hat{\mu}$ is the solution to (4.4) with initial condition $\hat{\mu}(0) = \xi$. The last equality (4.5) follows because $\hat{\mu}$ attains the minimum in (4.1) for each duration $[0, mT]$. □

The next lemma says that the optimal path above must end up in a specific invariant set for the dynamics given by the time-reversed McKean-Vlasov equation.

LEMMA 4.3. *Consider the trajectory $\hat{\mu}$ given by Lemma 4.2. Its ω-limit set, which is the set of its limit points as $t \to +\infty$ given by*

$$\Omega = \bigcap_{t>0} \overline{\{\hat{\mu}(t'), t' \geq t\}},$$

is positively invariant for the ODE

$$\dot{\hat{\mu}}(t) = -A^*_{\hat{\mu}(t)}\hat{\mu}(t), \quad t \geq 0. \tag{4.8}$$

PROOF. Take the $\hat{\mu}$ given by Lemma 4.2. It remains within $\mathcal{M}_1(\mathcal{Z})$, and satisfies the dynamics $\dot{\hat{\mu}}(t) = -\hat{L}(t)^*\hat{\mu}(t)$ for $t \in [0, +\infty)$ with initial condition $\hat{\mu}(0) = \xi$. Furthermore, (4.5) holds (for all $m \geq 1$). Since the integrand in (4.5) is nonnegative, $s(\hat{\mu}(mT))$ must decrease as m increases. But s is bounded between $[0, C_1(T)]$, and so there is an s^* such that $s(\hat{\mu}(mT)) \downarrow s^*$ as $m \uparrow +\infty$.

Consider any arbitrary subsequence of $(\hat{\mu}(mT), m \geq 1)$ and take a subsequential limit ξ'. On this subsequence, take a further subsequential limit of $(\hat{\mu}(mT + T), m \geq 1)$ and call it ν. Call the subsequence $(m_k, k \geq 1)$, and consider the paths

$$\mu^{(m_k)}(t) = \hat{\mu}(m_k T + T - t), \quad t \in [0, T]$$

which are of duration T and time reversals of fragments of $\hat{\mu}$. We thus have the subsequential convergence

$$(\mu^{(m_k)}(0), \mu^{m_k}(T)) \to (\nu, \xi') \text{ as } k \to +\infty. \tag{4.9}$$

Taking limits as $k \to +\infty$ in (4.5) and using the fact that $s(\mu^{(m_k)}(0))$ as well as $s(\mu^{(m_k)}(T))$ converge to s^* as $k \to +\infty$, the integral term, which is easily seen to be $S_{[0,T]}(\mu^{(m)}|\mu^{(m)}(0))$, satisfies

$$\limsup_{k \to +\infty} S_{[0,T]}(\mu^{(m_k)}|\mu^{(m_k)}(0)) = 0.$$

This fact and the nonnegativity of S_T imply

$$\lim_{k \to +\infty} S_T(\mu^{(m_k)}(T)|\mu^{(m_k)}(0)) = 0.$$

By the uniform continuity of S_T in both its arguments (Lemma 3.3) and by (4.9), we deduce that $S_T(\xi'|\nu) = 0$. But then the path that goes from ν to ξ' in time $[0, T]$ and attains $S_T(\xi'|\nu) = 0$ is the McKean-Vlasov path which is the solution to the dynamics

$$\dot{\mu}(t) = A^*_{\mu(t)}\mu(t), \ t \in [0, T]$$

with rate matrix $A_{\mu(t)}$, initial condition $\mu(0) = \nu$, and final condition $\mu(T) = \xi'$. But this implies that $\overline{\mu}(t) = \mu(T - t)$ satisfies (4.4) with $\hat{L}(t) = A_{\overline{\mu}(t)}$ and initial condition $\overline{\mu}(0) = \xi'$. It follows that Ω, the ω-limit set for $\hat{\mu}$, is contained in the ω-limit set for the dynamics (4.8) with initial condition $\hat{\mu}(0) = \xi$. This concludes the proof. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

4.1. *Invariant measure: Unique globally asymptotically stable equilibrium.* In this subsection, we consider the case when there is a unique globally asymptotically stable equilibrium ξ_0.

LEMMA 4.4. *If the McKean-Vlasov equation* $\dot{\mu}(t) = A^*_{\mu(t)}\mu(t)$ *has a unique globally asymptotically stable equilibrium* ξ_0*, then* $s(\xi_0) = 0$.

PROOF. Consider the McKean-Vlasov dynamics

$$\dot{\mu}(t) = A^*_{\mu(t)}\mu(t)$$

with initial condition $\mu(0) = \nu^*$. By our assumption that ξ_0 is the unique globally asymptotically stable equilibrium, $\mu(T) \to \xi_0$ as $T \to +\infty$. By the second remark following Theorem 3.2, the McKean-Vlasov path has zero cost and so $S_T(\mu(T)|\nu^*) = 0$ for each $T > 0$. By (4.1), we then have

$$s(\mu(T)) \leq s(\nu^*) + S_T(\mu(T)|\nu^*) = s(\nu^*).$$

Take limits as $T \to +\infty$ and use the lower semicontinuity of s to get

$$0 \leq s(\xi_0) \leq \liminf_{T\to+\infty} s(\mu(T)) \leq s(\nu^*) = 0,$$

whence $s(\xi_0) = 0$. □

The following lemma shows that the rate function for any subsequential large deviation principle is unique.

LEMMA 4.5. *If the McKean-Vlasov equation* $\dot{\mu}(t) = A^*_{\mu(t)}\mu(t)$ *has a unique globally asymptotically stable equilibrium* ξ_0*, then the solution to (4.1) and (4.5) is unique and is given by*

$$s(\xi) = \inf_{\hat{\mu}} \int_{[0,+\infty)} \left[\sum_{(i,j)\in\mathcal{E}} (\hat{\mu}(t)(i))\hat{\lambda}_{i,j}(\hat{\mu}(t))\tau^* \left(\frac{\hat{l}_{i,j}(t)}{\hat{\lambda}_{i,j}(\hat{\mu}(t))} - 1 \right) \right] dt \tag{4.10}$$

where the infimum is over all $\hat{\mu}$ *that are solutions to the dynamical system* $\dot{\hat{\mu}}(t) = -\hat{L}(t)^*\hat{\mu}(t)$ *for some family of rate matrices* $\hat{L}(\cdot)$*, with initial condition* $\hat{\mu}(0) = \xi$*, terminal condition* $\lim_{t\to+\infty} \hat{\mu}(t) = \xi_0$*, and* $\hat{\mu}(t) \in \mathcal{M}_1(\mathcal{Z})$ *for all* $t \geq 0$.

PROOF. The path given by Lemma 4.2 converges as shown in Lemma 4.3 to Ω, which is contained in an ω-limit set of the ODE (4.8). So Ω is connected and compact. Since the path $\hat{\mu}(\cdot)$ also stays completely within

$\mathcal{M}_1(\mathcal{Z})$ (viewed as a subset of $\mathbb{R}^r$), Ω is a subset of $\mathcal{M}_1(\mathcal{Z})$. Furthermore, Ω is invariant (both positively and negatively) to the dynamics defined by the ODE (4.8). But then, by our assumption that ξ_0 is the unique globally asymptotically stable equilibrium for the McKean-Vlasov dynamics

$$\dot{\mu}(t) = A^*_{\mu(t)}\mu(t), \tag{4.11}$$

we must have $\Omega = \{\xi_0\}$ because this is the only subset of $\mathcal{M}_1(\mathcal{Z})$ that is invariant to both of the dynamics in (4.8) and (4.11). Letting $m \to +\infty$ in (4.5), it follows by lower semicontinuity of s that

$$s(\xi) \geq s(\xi_0) + \int_{[0,+\infty)} \Big[\sum_{(i,j)\in\mathcal{E}} (\hat{\mu}(t)(i))\hat{\lambda}_{i,j}(\hat{\mu}(t))\tau^* \left(\frac{\hat{l}_{i,j}(t)}{\hat{\lambda}_{i,j}(\hat{\mu}(t))} - 1 \right) \Big]\, dt. \tag{4.12}$$

By Lemma 4.4, $s(\xi_0) = 0$, and thus

$$s(\xi) \geq \int_{[0,+\infty)} \Big[\sum_{(i,j)\in\mathcal{E}} (\hat{\mu}(t)(i))\hat{\lambda}_{i,j}(\hat{\mu}(t))\tau^* \left(\frac{\hat{l}_{i,j}(t)}{\hat{\lambda}_{i,j}(\hat{\mu}(t))} - 1 \right) \Big]\, dt$$

for the special path $\hat{\mu}$. This establishes that $s(\xi)$ is at least the right-hand side of (4.10). By an application of (4.1) with $\nu = \xi_0$, it is obvious that $s(\xi)$ is upper bounded by the right-hand side of (4.10), whence equality holds in (4.10) and the uniqueness of $s(\xi)$ follows. □

We are now ready to prove Theorem 2.1.

PROOF OF THEOREM 2.1. Take any arbitrary sequence of natural numbers growing to $+\infty$. By Lemma 4.1, there is a subsequence that satisfies the large deviation principle with rate function s such that (4.1) holds. Given our assumption that ξ_0 is the unique globally asymptotically stable equilibrium for the McKean-Vlasov equation, we have $s \geq 0$, and $s(\xi_0) = 0$ by Lemma 4.4. By Lemma 4.5, s is uniquely specified by (4.10), which is the same as (2.7). Thus every sequence contains a further subsequence $(N_k, k \geq 1)$ such that $(\wp^{(N_k)}, k \geq 1)$ satisfies the large deviation principle with speed N_k and the same rate function s specified by (2.7). By [13, Ex. 4.4.15(a)-(b), pp. 147-148], it follows that $(\wp^{(N)}, N \geq 1)$ satisfies the large deviation principle with speed N and rate function given by (2.7). □

4.2. *Invariant measure: The general case.* We now treat the general case under assumption (**B**). As noted earlier, under the hypothesis of Theorem 2.1, assumption (**B**) holds with $l = 1$, $K_1 = \{\xi_0\}$, and moreover $s(\xi_0) = 0$ by Lemma 4.4. We generalize Lemma 4.4 in Lemma 4.8 by first proving the following lemma. The generalization of Lemma 4.5 is Lemma 4.7.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA 4.6. *Under assumption* (**B**), *the rate function* s *satisfies the following:*

- *There exists* $\xi_0 \in K_{i_0}$ *for some* $i_0 = 1, 2, \ldots, l$ *that satisfies* $s(\xi_0) = 0$.
- *There exist nonnegative real numbers* $s_1, s_2, \ldots, s_l$ *such that* $\xi \in K_i$ *implies* $s(\xi) = s_i$.

PROOF. The first statement immediately follows from the steps in the proof of Lemma 4.4: ξ_0 is now some element in the ω-limit set for the McKean-Vlasov dynamics with initial condition ν^* satisfying $s(\nu^*) = 0$.

For the second statement, let $\nu, \xi \in K_i$. Fix $\varepsilon > 0$. Since $\nu \sim \xi$, there exists $T > 0$ such that $S_T(\xi|\nu) \leq \varepsilon$. Using (4.1), we get

$$s(\xi) \leq s(\nu) + S_T(\xi|\nu) \leq s(\nu) + \varepsilon.$$

Reversing the role of ξ and ν, we get $s(\nu) \leq s(\xi) + \varepsilon$, whence $|s(\xi) - s(\nu)| \leq \varepsilon$. Since ε was arbitrary, we must have $s(\nu) = s(\xi)$. So all points in the compact set K_i take the same value. The second statement follows. □

We now argue that the function s that satisfies (4.1), (4.5), the condition $s \geq 0$, and the condition $\min_\nu s(\nu) = 0$ is indeed unique. We do this in two steps via the following lemmas.

LEMMA 4.7. *Assume that* (**B**) *holds. Let* $s_1, s_2, \ldots, s_l$ *be specified as the values on the compact sets* $K_1, K_2, \ldots, K_l$. *Let* $s_{l'} \geq 0$ *for* $1 \leq l' \leq l$ *and let* $\min\{s_1, s_2, \ldots, s_l\} = 0$. *Then the solution to (4.1) and (4.5) is unique and is given by*

(4.13)
$$s(\xi) = \inf_{l'} \inf_{\hat{\mu}} \left[s_{l'} + \int_0^{+\infty} \Big[\sum_{(i,j) \in \mathcal{E}} (\hat{\mu}(t)(i)) \hat{\lambda}_{i,j}(\hat{\mu}(t)) \tau^* \left(\frac{\hat{l}_{i,j}(t)}{\hat{\lambda}_{i,j}(\hat{\mu}(t))} - 1 \right) \Big] \, dt \right]$$

where the second infimum is over all $\hat{\mu}$ *that are solutions to the dynamical system* $\dot{\hat{\mu}}(t) = -\hat{L}(t)^* \hat{\mu}(t)$ *for some family of rate matrices* $\hat{L}(\cdot)$, *with initial condition* $\hat{\mu}(0) = \xi$, *terminal condition* $\hat{\mu}(t) \to K_{l'}$ *as* $t \to +\infty$, *and* $\hat{\mu}(t) \in \mathcal{M}_1(\mathcal{Z})$ *for all* $t \geq 0$.

PROOF. The same steps of the proof of Lemma 4.5 apply with the following modifications. $\Omega = \{\xi_0\}$ gets replaced by $\Omega \subset K_{l'}$ for some l'. Consequently, in the lower bound in (4.12), $s(\xi_0)$ gets replaced by $s_{l'}$. □

Recall that we defined $s_1, s_2, \ldots, s_l$ in (2.13). We now assert the following.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA 4.8. *The rate function s has values $s_1, s_2, \ldots, s_l$ given by (2.13).*

PROOF. Immediate from Theorem A.1 in Appendix. □

REMARK. In order to obtain these values, one has to go beyond the ODE method. One has to consider the empirical measure Markov process sampled at hitting times of neighborhoods of the compact sets. This is done in Freidlin and Wentzell [21, Ch. 6] for diffusions on a compact manifold, with V satisfying a Lipschitz property. Thanks to Corollary 3.1, Lemma 3.2, Lemma 3.3, the same program can be carried out with straightforward modifications to account for the fact that we have to handle jumps and the fact that the minimum cost function $V(\cdot|\cdot)$ satisfies only the uniform continuity property. The appendix provides the necessary verification.

With this disambiguation for the values of s at the compact sets K_i, we now ready to finish the proof of Theorem 2.2.

PROOF OF THEOREM 2.2. Same as that of Theorem 2.1, but with the use of Lemmas 4.7 and 4.8. □

The following sections complete the proofs of some assertions of Section 3.

5. Proof of Theorem 3.1. The proof is based on a generalization of Sanov's theorem due to Dawson and Gärtner [11], the Girsanov transformation, and the Laplace-Varadhan principle. We proceed through a sequence of lemmas.

5.1. *The noninteracting case.* Consider first the noninteracting case.

LEMMA 5.1. *Suppose that the initial conditions $\nu_N \to \nu$ weakly. Then the sequence $(P^{o,(N)}_{\nu_N}, N \geq 1)$ satisfies the large deviation principle in $\mathcal{M}_{1,\varphi}(\mathcal{X})$, endowed with the weak* topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$, with speed N and good rate function $J(Q)$ given by (3.9).*

PROOF. The family $\{P_z, z \in \mathcal{Z}\}$ is clearly a subset of $\mathcal{M}_{1,\varphi}(\mathcal{X})$ since the transition rate from any state to any of the other (at most r) states is upper bounded by 1. The family $\{P_z, z \in \mathcal{Z}\}$ is also Feller continuous in the discrete topology on $\mathcal{Z}$. Since $\nu_N \to \nu$, by Dawson and Gärtner's [11, Th. 3.5] which is a generalization of Sanov's theorem, we have that $(P^{o,(N)}_{\nu_N}, N \geq 1)$ satisfies the large deviation principle in the weak* topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$ with speed N and good rate function $J(Q)$ given by (3.9). □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Our next lemma states that $\mathcal{M}_{1,\varphi}(\mathcal{X})$ contains all the probability measures of interest to us.

LEMMA 5.2. *If $J(Q) < +\infty$ then (1) $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$ and (2) $Q \circ \pi_0^{-1} = \nu$.*

PROOF. Observe that $||\varphi||_\varphi \leq 1$, and the topology on $\mathcal{X}$ is chosen so that φ is continuous. Hence $\varphi \in C_\varphi(\mathcal{X})$. Recalling the expression for $J(Q)$ in (3.9), we get that $J(Q) < +\infty$ implies

$$\int_{\mathcal{X}} \varphi \, dQ - \sum_{z \in \mathcal{Z}} \nu(z) \log \int_{\mathcal{X}} e^{\varphi} \, dP_z < +\infty. \tag{5.1}$$

Since the transition rates for P_z are upper bounded by 1, and there are at most r possibilities for jumps from any state, φ is stochastically dominated by a Poisson random variable with parameter rT. Consequently $1 \leq \int_{\mathcal{X}} e^{\varphi} \, dP_z < +\infty$ for each $z \in \mathcal{Z}$, and it follows from (5.1) that $\int_{\mathcal{X}} \varphi \, dQ < +\infty$, and so $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$.

To prove the second conclusion, suppose $\nu_Q = Q \circ \pi_0^{-1} \neq \nu$. Consider bounded functions $f(x) = f_0(\pi_0(x))$ that depend on x only through the initial condition. Since $\nu_Q \neq \nu$, there exists a function f of the above form that also satisfies $\sum_z f_0(z)\nu_Q(z) - \sum_z f_0(z)\nu(z) \neq 0$. By flipping the sign of f if necessary and by scaling, we may assume that $\sum_z f_0(z)\nu_Q(z) - \sum_z f_0(z)\nu(z) = a$ for an arbitrary $a \in (0, +\infty)$. This f is bounded continuous and hence is in $C_\varphi(\mathcal{X})$. A simple calculation yields

$$\int_{\mathcal{X}} f \, dQ - \sum_{z \in \mathcal{Z}} \nu(z) \log \int_{\mathcal{X}} e^{f} \, dP_z = \sum_z f_0(z)\nu_Q(z) - \sum_z f_0(z)\nu(z) = a.$$

Since $a > 0$ was arbitrary, $J(Q) = +\infty$, and the second part is proved by contraposition. □

We next get an alternative expression for $J(Q)$ as a supremum over the more familiar space $C_b(\mathcal{X})$ of bounded continuous functions on $\mathcal{X}$.

LEMMA 5.3. *For each $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$, $J(Q)$ defined in (3.9) admits the alternative characterization*

$$J(Q) = \sup_{f \in C_b(\mathcal{X})} \left[\int_{\mathcal{X}} f \, dQ - \sum_{z \in \mathcal{Z}} \nu(z) \log \int_{\mathcal{X}} e^{f} \, dP_z \right]. \tag{5.2}$$

PROOF. In (5.2), the supremum is taken over $C_b(\mathcal{X})$, the set of bounded continuous functions on $\mathcal{X}$, while in (3.9) the supremum is over $C_\varphi(\mathcal{X})$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Fix $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$. Since $C_b(\mathcal{X}) \subset C_\varphi(\mathcal{X})$, $J(Q)$ defined by (3.9) is at least the right-hand side of (5.2). For the other direction, let $f \in C_\varphi(\mathcal{X})$, and consider the truncations f_n of the function f to $[-n, n]$. Clearly $\{f_n\} \subset C_b(\mathcal{X})$, and $f_n \to f$ a.e.-$[Q]$ and a.e.-$[P_z]$. Since we also have $\int_{\mathcal{X}} f \, dQ < +\infty$ and $\int_{\mathcal{X}} e^f \, dP_z < +\infty$, both of which can be easily checked using $||f||_\varphi < +\infty$, an application of the Lebesgue dominated convergence theorem yields

$$\begin{aligned}\lim_{n\to+\infty} \left[\int_{\mathcal{X}} f_n \, dQ - \sum_{z\in\mathcal{Z}} \nu(z) \log \int_{\mathcal{X}} e^{f_n} \, dP_z\right] \\ = \int_{\mathcal{X}} f \, dQ - \sum_{z\in\mathcal{Z}} \nu(z) \log \int_{\mathcal{X}} e^f \, dP_z.\end{aligned}$$

Since $f \in C_\varphi(\mathcal{X})$ was arbitrary, $J(Q)$ in (3.9) is at most the right-hand side of (5.2). □

We now characterize $J(Q)$ further. We begin by getting a lower bound.

LEMMA 5.4. *Let Q be such that $Q \circ \pi_0^{-1} = \nu$. Define P as in (3.6). We then have $H(Q|P) \le J(Q)$.*

PROOF. For any $f \in C_b(\mathcal{X})$ and with P as defined in (3.6), Jensen's inequality yields

$$\int_{\mathcal{X}} f \, dQ - \sum_z \nu(z) \log \int_{\mathcal{X}} e^f \, dP_z \ge \int_{\mathcal{X}} f \, dQ - \log \int_{\mathcal{X}} e^f \, dP.$$

By Lemma 5.3, $J(Q)$ is the supremum of the left-hand side over all $f \in C_b(\mathcal{X})$. By the variational formula for relative entropy ([13, Lem. 6.2.13]), $H(Q|P)$ is the supremum of the right-hand side over all $f \in C_b(\mathcal{X})$. This establishes the inequality. □

We next show that $J(Q)$ is upper bounded by another relative entropy. To do this, let us introduce the Polish space $(\hat{\mathcal{X}}, \hat{d})$ where $\hat{\mathcal{X}} = \mathcal{Z} \times \mathcal{X}$ and the metric $\hat{d}$ is

$$\hat{d}((i,x),(j,y)) = \mathbf{1}_{\{i\neq j\}} + d(x,y),$$

where d is the metric on $\mathcal{X}$. The first component of $\hat{\mathcal{X}}$ shall denote the initial condition. For two measures $\hat{R}_1$ and $\hat{R}_2$ on $\hat{\mathcal{X}}$, let $H(\hat{R}_1|\hat{R}_2)$ be the relative entropy of $\hat{R}_1$ with respect to $\hat{R}_2$.

For a fixed $\nu \in \mathcal{M}_1(\mathcal{Z})$, let us now define $\hat{P}$ as

$$d\hat{P}(z,x) = \nu(z) dP_z(x). \tag{5.3}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

The push forward of $\hat{P}$ under the projection mapping $(z,x) \in \hat{\mathcal{X}} \mapsto x \in \mathcal{X}$ is clearly the P defined in (3.6).

Let Q be such that $Q \circ \pi_0^{-1} = \nu$. We then define $\hat{Q}$ as

$$d\hat{Q}(z,x) = dQ(x)\mathbf{1}_{\{\pi_0(x)\}}(z). \tag{5.4}$$

Observing that $\hat{\mathcal{X}}$ is Polish and that the push forward of $\hat{Q}$ under the projection $(z,x) \mapsto z$ is ν, it follows that there is a regular conditional probability measure Q_z satisfying

$$d\hat{Q}(z,x) = \nu(z)dQ_z(x). \tag{5.5}$$

Putting (5.4) and (5.5) together and summing over z, we obtain that the second marginal of $\hat{Q}$ is

$$dQ(x) = \sum_z dQ(x)\mathbf{1}_{\{\pi_0(x)\}}(z) = \sum_z \nu(z)dQ_z(x). \tag{5.6}$$

Since both $\hat{Q}$ and $\hat{P}$ have the same first marginal ν, the decomposition result for relative entropy [13, Th. D.8] gives

$$H(\hat{Q}|\hat{P}) = \begin{cases} \displaystyle\sum_{z\in\mathcal{Z}} \nu(z)H(Q_z|P_z) & \text{if } Q_z \ll P_z \text{ a.e.-}[\nu] \\ +\infty & \text{otherwise.} \end{cases} \tag{5.7}$$

With these preliminaries, we are now ready to upper bound $J(Q)$.

LEMMA 5.5. *Let Q be such that $Q \circ \pi_0^{-1} = \nu$. With $\hat{P}$ and $\hat{Q}$ as above, we have $J(Q) \le H(\hat{Q}|\hat{P})$.*

PROOF. For $f \in C_b(\mathcal{X})$, by (5.6), we have $\int_{\mathcal{X}} f\ dQ = \sum_z \nu(z) \int_{\mathcal{X}} f\ dQ_z$, and so

$$\begin{aligned} &\int_{\mathcal{X}} f\ dQ - \sum_z \nu(z) \log \int_{\mathcal{X}} e^f\ dP_z \\ &= \sum_z \nu(z) \left[\int_{\mathcal{X}} f\ dQ_z - \log \int_{\mathcal{X}} e^f\ dP_z \right] \\ &\le \sum_z \nu(z) \left(\sup_{f\in C_b(\mathcal{X})} \left[\int_{\mathcal{X}} f\ dQ_z - \log \int_{\mathcal{X}} e^f\ dP_z \right] \right) \\ &= H(\hat{Q}|\hat{P}), \end{aligned}$$

where the last equality follows from the observation that the term within parenthesis in the immediately preceding inequality is the variational representation for the relative entropy $H(Q_z|P_z)$, and from (5.7). Take supremum over all $f \in C_b(\mathcal{X})$ and use Lemma 5.3 to deduce that $J(Q)$ is upper bounded by $H(\hat{Q}|\hat{P})$. □

LEMMA 5.6. *Let $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$. We then have*

$$J(Q) = \begin{cases} H(Q|P), & \text{if } Q \circ \pi_0^{-1} = \nu \\ +\infty & \text{otherwise.} \end{cases}$$

PROOF. If $Q \circ \pi_0^{-1} \neq \nu$, by Lemma 5.2, we have $J(Q) = +\infty$. So assume $Q \circ \pi_0^{-1} = \nu$. By Lemmas 5.4 and 5.5, we get

$$H(Q|P) \leq J(Q) \leq H(\hat{Q}|\hat{P})$$

where $\hat{P}$ and $\hat{Q}$ are defined in (5.3) and (5.4), respectively. The second marginal of $\hat{P}$ is $P = \sum_z \nu(z) P_z$. From the definition of $\hat{P}$ it is clear that $d\hat{P}(z,x) = dP(x)\mathbf{1}_{\{\pi_0(x)\}}(z)$, so that the regular conditional probability measures of both $\hat{P}$ and $\hat{Q}$, given the second component x, are the same, that is, $d\hat{Q}(z|x) = d\hat{P}(z|x) = \mathbf{1}_{\{\pi_0(x)\}}(z)$ a.e.-$[Q]$. In particular, $H(\hat{Q}(\cdot|x)|\hat{P}(\cdot|x)) = 0$ a.e.-$[Q]$. By the decomposition result for relative entropy [13, Th. D.8], we get

$$\begin{aligned} H(\hat{Q}|\hat{P}) &= H(Q|P) + \int_{\mathcal{X}} H(\hat{Q}(\cdot|x)|\hat{P}(\cdot|x))\, dQ(x) \\ &= H(Q|P). \end{aligned}$$

The lemma follows. □

5.2. *Continuity of the function $h(Q)$.* We now proceed to address some preliminaries required for the interacting case. The Radon-Nikodym derivative (3.4) of $P_{\nu_N}^{(N)}$ with respect to $P_{\nu_N}^{o,(N)}$ is

$$\frac{dP_{\nu_N}^{(N)}}{dP_{\nu_N}^{o,(N)}}(Q) = \exp\{Nh(Q)\}.$$

We now study the continuity property of $h(Q)$. Towards this, we first establish a regularity property for all Q with $J(Q) < +\infty$. We then appeal to results of [24] to establish the continuity of $h(Q)$ when $J(Q) < +\infty$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA 5.7. *Let $J(Q) < +\infty$ and suppose that the random variable X is distributed according to Q. Then*

$$\sup_{t\in[0,T]} \mathbb{E}\left[\sup_{u\in[t-\alpha,t+\alpha]\cap[0,T]} \{\mathbf{1}_{\{X_u \neq X_{u-}\}}\}\right] \to 0 \text{ as } \alpha \downarrow 0. \tag{5.8}$$

PROOF. A proof for the case when $Q \circ \pi_0^{-1} = \delta_{z_0}$ for some fixed z_0 can be found in [24, eqn. (2.14), p. 309-310]. Our argument below is a simple modification and relies only on what we have thus far established for $J(Q)$.

Let $K = \{x \in \mathcal{X} : \exists u \in [t-\alpha, t+\alpha] \cap [0,T] \text{ satisfying } x_u \neq x_{u-}\}$. Since $J(Q) < +\infty$, it follows that $Q \ll P$. We may therefore write

$$\begin{aligned} \mathbb{E}\left[\sup_{u\in[t-\alpha,t+\alpha]\cap[0,T]} \{\mathbf{1}_{\{X_u \neq X_{u-}\}}\}\right] &= Q(K) \\ &= \int_{\mathcal{X}} \left(\frac{dQ}{dP}\right) \mathbf{1}_K \, dP \\ &\leq \left\|\frac{dQ}{dP}\right\|_{\tau^*,P} \|\mathbf{1}_K\|_{\tau,P} \end{aligned} \tag{5.9}$$

where $||f||_{\tau^*,P}$ is the Orlicz norm

$$||f||_{\tau^*,P} = \inf\left\{a > 0 : \int_{\mathcal{X}} \tau^*\left(\frac{|f(x)|}{a}\right) dP(x) \leq 1\right\}$$

(with respect to the function τ^* and measure P), $||g||_{\tau,P}$ is a similarly defined Orlicz norm with respect to the function τ and measure P, and the inequality in (5.9) is the Hölder inequality in Orlicz spaces. See the Appendix in [24] for a summary of key results on Orlicz spaces.

The lemma's proof will be complete if we can show that $\left\|\frac{dQ}{dP}\right\|_{\tau^*,P}$ is bounded, and $\|\mathbf{1}_K\|_{\tau,P}$ vanishes as $\alpha \downarrow 0$. We proceed to justify these claims.

$J(Q) < +\infty$ implies $\left\|\frac{dQ}{dP}\right\|_{\tau^*,P} < +\infty$. Indeed, since

$$\lim_{u\to+\infty} \frac{\tau^*(u)}{u \log u} = 1,$$

choose a large enough $u_0 > 0$ such that $\tau^*(u) \leq 2u \log u$ for $u \geq u_0$. Also observe that $\tau^*(u)$ is increasing in u and $u \log u \geq -e^{-1}$ for $u \geq 0$. Thus with $f = dQ/dP$,

$$\begin{aligned} \int_{\mathcal{X}} \tau^*(f) \, dP &\leq \tau^*(u_0) + \int_{\{x\in\mathcal{X}: f(x)\geq u_0\}} 2f \log f \, dP \\ &\leq \tau^*(u_0) + 2J(Q) + 2e^{-1} \\ &< +\infty. \end{aligned} \tag{5.10}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Since τ^* is convex and $\tau^*(0) = 0$, Jensen's inequality yields

$$\tau^*(f/a) \leq \tau^*(f)/a \text{ for } a \geq 1.$$

This fact in conjunction with (5.10) implies that $||f||_{\tau^*,P} < +\infty$.

Now consider $\|\mathbf{1}_K\|_{\tau,P}$. Since $\tau(0) = 0$, we get $\tau((\mathbf{1}_K)/a) = \tau(1/a)\mathbf{1}_K$, and so

$$\int_{\mathcal{X}} \tau\left(\frac{\mathbf{1}_K}{a}\right)\, dP = \tau(1/a) \int_{\mathcal{X}} \mathbf{1}_K\, dP = \tau(1/a)P(K).$$

Under P, the transition rates are upper bounded by 1. Moreover, there are at most r possible next states. Since K is the event that there is a transition in $[t-\alpha, t+\alpha] \cap [0,T]$, it follows that $P(K) \leq 2\alpha r$. From its definition, the Orlicz norm is the smallest positive a such that $\tau(1/a)P(K) \leq 1$, and so

$$\|\mathbf{1}_K\|_{\tau,P} = \frac{1}{\tau^{-1}(1/P(K))} \leq \frac{1}{\tau^{-1}(1/(2\alpha r))}.$$

where the equality in the above chain holds because $\tau(u)$ is increasing in u for $u \geq 0$, and the inequality holds because of the same property for $\tau^{-1}(u)$. This last upper bound vanishes as $\alpha \downarrow 0$. □

The following lemma implies Lemma 3.1 as a corollary.

LEMMA 5.8. *Consider $\mathcal{M}_1(D([0,T],\mathcal{Z}))$ endowed with the topology of weak convergence and $D([0,T],\mathcal{M}_1(\mathcal{Z}))$ endowed with the topology induced by the metric ρ_T defined in (3.11). Then the mapping*

$$\pi : \mathcal{M}_1(D([0,T],\mathcal{Z})) \to D([0,T],\mathcal{M}_1(\mathcal{Z}))$$

is continuous at each $Q \in \mathcal{M}_1(D([0,T],\mathcal{Z}))$ where $J(Q) < +\infty$.

PROOF. Note that, by Lemma 5.2, $J(Q) < +\infty$ implies $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$. The statement of the above Lemma is the same as [24, Lem. 2.8]. The only difference is that our representation for $J(Q)$ differs in order to handle nonchaotic initial conditions and allows $Q \circ \pi_0^{-1}$ to be any measure in $\mathcal{M}_1(\mathcal{Z})$.

The proof of [24, Lem. 2.8] holds verbatim if we can establish (5.8) (which is the same as [24, eqn. (2.14)]) for the more general case under consideration. This is done in Lemma 5.7. □

This is an appropriate location to include the proof of Lemma 3.1, which is a corollary to the above Lemma.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

PROOF OF LEMMA 3.1. The first part is a corollary to Lemma 5.8. Indeed, Lemma 5.8 shows that π is continuous under the coarser topology of weak convergence of probability measures metrized by d_{Sko}. Since the natural embedding of $\mathcal{X}$ into $D([0,T],\mathcal{Z})$ (with topology induced by d_{Sko}) is continuous, it immediately follows that π is a continuous mapping under the finer topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$.

To see the second part, fix a Q such that $J(Q) < +\infty$, a $t \in [0,T]$, and consider a sequence $Q_N \to Q$. By the first part, we have $\pi(Q_N) \to \pi(Q)$, which is the same as saying $\rho_T(\pi(Q_N), \pi(Q)) \to 0$. But then

$$\rho_0(\pi_t(Q_N), \pi_t(Q)) \le \rho_T(\pi(Q_N), \pi(Q)) \to 0$$

establishes the continuity of π_t. □

We now come to the continuity of the function $h(Q)$.

LEMMA 5.9. *Consider the space $\mathcal{M}_{1,\varphi}(\mathcal{X})$ endowed with the weak* topology $\sigma(\mathcal{M}_{1,\varphi}(\mathcal{X}), C_\varphi(\mathcal{X}))$. The function $h : \mathcal{M}_{1,\varphi}(\mathcal{X}) \to \mathbb{R}$ defined in (3.5) is continuous at every Q where $J(Q) < +\infty$.*

PROOF. The statement is the same as [24, Lem. 2.9]. The same proof applies. That proof requires continuity of π, which is now established in Lemma 5.8 under assumptions (**A1**)-(**A3**). □

5.3. *The interacting case.* We now address the interacting case.

PROOF OF THEOREM 3.1. Recall the statement of Theorem 3.1. We are now given that the sequence of initial empirical measures $\nu_N \to \nu$ weakly. By Lemma 5.1, $(P^{o,(N)}_{\nu_N}, N \ge 1)$ satisfies the large deviation principle in the topological space $\mathcal{M}_{1,\varphi}(\mathcal{X})$ with rate function $J(Q)$. By Lemma 5.2 and Lemma 5.9, h is continuous on the set $\{Q \in \mathcal{M}_{1,\varphi}(\mathcal{X}) \mid J(Q) < +\infty\}$. Furthermore, by [24, Lem. 2.10], for every $\alpha > 0$, we have

$$\limsup_{N\to+\infty} \frac{1}{N} \log \int_{\mathcal{M}_{1,\varphi}(\mathcal{X})} e^{N\alpha|h|}\, dP^{o,(N)}_{\nu_N} < +\infty.$$

Using the Laplace-Varadhan principle, see [24, Prop. 2.5], we can draw two conclusions. The first conclusion is that

$$\frac{1}{N} \log \int_{\mathcal{M}_{1,\varphi}(\mathcal{X})} e^{Nh}\, dP^{o,(N)}_{\nu_N} \to \sup_{Q' \in \mathcal{M}_{1,\varphi}(\mathcal{X})} [h(Q') - J(Q')] \tag{5.11}$$

as $N \to +\infty$. From (3.4), we have

$$e^{Nh} dP^{o,(N)}_{\nu_N} = dP^{(N)}_{\nu_N},$$

a probability measure. The left-hand side in (5.11) is therefore always 0, and so $\sup_{Q' \in \mathcal{M}_{1,\varphi}(\mathcal{X})}[h(Q') - J(Q')] = 0$. The second conclusion is that $(P^{(N)}_{\nu_N}, N \geq 1)$ satisfies the large deviation principle in the topological space $\mathcal{M}_{1,\varphi}(\mathcal{X})$ with good rate function

$$I(Q) = J(Q) - h(Q) - \inf_{Q'}[J(Q') - h(Q')] = J(Q) - h(Q)$$

where the last equality holds because the infimum above is 0 by the first conclusion. This concludes the proof of the first part of Theorem 3.1.

We now show (3.10). By assumption (**A1**)-(**A3**), it is easy to see that there exists a constant K such that $|h(Q)| \leq K(1 + \int_{\mathcal{X}} \varphi \, dQ)$ so that if $Q \in \mathcal{M}_{1,\varphi}(\mathcal{X})$ then $|h(Q)| < +\infty$. By Lemma 5.6, if either $Q \circ \pi_0^{-1} \neq \nu$ or Q is not absolutely continuous with respect to P, then $J(Q) = +\infty$, and by the finiteness of $h(Q)$, we have $I(Q) = J(Q) - h(Q) = +\infty$. We may therefore assume $Q \circ \pi_0^{-1} = \nu$ and $Q \ll P$, whence, by Lemma 5.6 once again, $J(Q) = H(Q|P) = \int dQ \ \log(dQ/dP)$. It therefore suffices to argue that

$$Q \circ \pi_0^{-1} = \nu \text{ and } Q \ll P \quad \Rightarrow \quad H(Q|P) - h(Q) = H(Q|P(\pi(Q))).$$

Let $\mu = \pi(Q)$ for convenience. Observe that the density

$$\frac{dP_z(\mu)}{dP_z}(\cdot) = \exp\{h_1(\cdot, \mu)\}$$

in (3.2) does not depend on z. It follows that the density of the mixture distribution $P(\mu)$ in (3.7) with respect to the mixture P in (3.6) is

$$\frac{dP(\mu)}{dP}(x) = \exp\{h_1(x, \mu)\}.$$

Using this in (3.5), we get

$$h(Q) = \int_{D([0,T],\mathcal{Z})} dQ \ \log \frac{dP(\mu)}{dP},$$

from which

$$\begin{aligned} H(Q|P) - h(Q) &= \int_{D([0,T],\mathcal{Z})} dQ \ \log \frac{dQ}{dP} - \int_{D([0,T],\mathcal{Z})} dQ \ \log \frac{dP(\mu)}{dP} \\ &= \int_{D([0,T],\mathcal{Z})} dQ \ \log \frac{dQ}{dP(\mu)} \\ &= H(Q|P(\mu)) \end{aligned}$$

follows. This concludes the proof. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

6. Proof of Lemma 3.2. We address the first bullet. For ease of exposition, let us for now allow all possible transitions and ignore the constraint that only $\mathcal{E}$ transitions are allowed. Consider the constant velocity path

$$\mu(t) = \left(1 - \frac{t}{T}\right)\nu + \frac{t}{T}\xi, \ t \in [0,T] \tag{6.1}$$

for which

$$\dot{\mu}(t) = \frac{\xi - \nu}{T}, \ t \in [0,T].$$

There is flow out of i if $\xi(i) < \nu(i)$, and flow into i otherwise. We now construct a rate matrix $L(t)$ with entries $l_{i,j}(t)$ that ensure the traversal of this constant velocity path.

Since there is conservation of mass $\sum_z \xi(z) = \sum_z \nu(z)$, we can construct mass transport parameters $\{g_{i,j}\}$ such that for an i with $\nu(i) > \xi(i)$ and a j with $\nu(j) < \xi(j)$, the quantity $g_{i,j}$ is the fraction of the excess $\nu(i) - \xi(i)$ that goes from i to j. In particular, $\{g_{i,j}\}$ satisfies

$$\begin{aligned} g_{i,j} &\in [0,1] \quad \text{for all } i,j \in \mathcal{Z} \\ g_{i,j} &= 0 \quad \text{if } \nu(i) \le \xi(i) \text{ or } \nu(j) \ge \xi(j) \qquad (6.2) \\ \sum_{j:\nu(j)<\xi(j)} g_{i,j} &= 1 \quad \text{if } \nu(i) > \xi(i) \qquad (6.3) \end{aligned}$$

and finally

$$\sum_{i:\nu(i)>\xi(i)} [\nu(i) - \xi(i)] g_{i,j} = \xi(j) - \nu(j) \quad \text{if } \nu(j) < \xi(j). \tag{6.4}$$

Equation (6.3) says mass is not destroyed and (6.4) says new mass is not created (all mass entering j must come from i's with excesses).

Define the diagonal elements of the transition rate matrix $L(t)$ to be

$$l_{j,j}(t) = \begin{cases} \frac{-(\nu(j)-\xi(j))}{T(\mu(t)(j))} & \text{if } j \text{ satisfies } \nu(j) > \xi(j) \\ 0 & \text{otherwise.} \end{cases} \tag{6.5}$$

Now define the off-diagonal elements of $L(t)$ to be

$$l_{j,i}(t) = \begin{cases} 0 & \text{if } \nu(j) \le \xi(j), i \in \mathcal{Z} \\ -l_{j,j}(t) g_{j,i} & \text{if } \nu(j) > \xi(j) \text{ and } i \ne j. \end{cases} \tag{6.6}$$

We next claim that $\dot{\mu}(t) = L(t)^* \mu(t)$. Indeed, for i such that $\nu(i) \ge \xi(i)$, we have

$$\begin{aligned} &(L(t)^* \mu(t))(i) \\ &\quad = \sum_j (\mu(t)(j))\, l_{j,i}(t) \end{aligned}$$

$$
\begin{aligned}
&= (\mu(t)(i))\ l_{i,i}(t) + \sum_{j:j\neq i,\nu(j)\leq\xi(j)} (\mu(t)(j))\ l_{j,i}(t) + \sum_{j:j\neq i,\nu(j)>\xi(j)} (\mu(t)(j))\ l_{j,i}(t) \\
&\overset{(a)}{=} (\mu(t)(i))\ l_{i,i}(t) + 0 + 0 \\
&\overset{(b)}{=} \frac{\xi(i)-\nu(i)}{T}.
\end{aligned}
$$

In the above sequence of equalities, the second term in (a) vanished because $\nu(j) \leq \xi(j)$ implies $l_{j,i}(t) = 0$ (see (6.6)); the third term vanished because, by (6.2) and noticing that i is the second argument, $\nu(i) \geq \xi(i)$ implies $g_{j,i} = 0$ which in turn implies $l_{j,i}(t) = 0$ again by (6.6). Lastly, (b) follows from (6.5).

For i such that $\nu(i) < \xi(i)$, we have

$$
\begin{aligned}
&(L(t)^*\mu(t))\,(i) \\
&= \sum_j (\mu(t)(j))\ l_{j,i}(t) \\
&= (\mu(t)(i))\ l_{i,i}(t) + \sum_{j:j\neq i,\nu(j)\leq\xi(j)} (\mu(t)(j))\ l_{j,i}(t) + \sum_{j:j\neq i,\nu(j)>\xi(j)} (\mu(t)(j))\ l_{j,i}(t) \\
&\overset{(a)}{=} 0 + 0 + \sum_{j:j\neq i,\nu(j)>\xi(j)} (\mu(t)(j))\ (-l_{j,j}(t)g_{j,i}) \\
&\overset{(b)}{=} \frac{1}{T} \sum_{j:j\neq i,\nu(j)>\xi(j)} (\nu(j)-\xi(j))g_{j,i} \\
&\overset{(c)}{=} \frac{\xi(i)-\nu(i)}{T}.
\end{aligned}
$$

In the above sequence of equalities, (a) follows from (6.5) and (6.6). Equation (b) follows from (6.5), and (c) follows from (6.4). The above arguments establish $\dot{\mu}(t) = L(t)^*\mu(t)$.

Let us now evaluate the difficulty $S_{[0,T]}(\mu|\nu)$ of passage near this constant velocity path μ. If we show that the integral in the right-hand side of (2.8) is finite, by Theorem 3.2, $S_{[0,T]}(\mu|\nu)$ equals this integral. Observe that $l_{i,j}(t)$ is not bounded if one of $\mu(0)(i) = \nu(i)$ or $\mu(T)(i) = \xi(i)$ equals 0, and so we do have some work to do.

The right-hand side of (2.8) can be expanded to be

$$
\begin{aligned}
\int_{[0,T]} \Big[\sum_{i,j:j\neq i} \Big(&(\mu(t)(i))\ l_{i,j}(t) \log\Big(\frac{l_{i,j}(t)}{\lambda_{i,j}(\mu(t))}\Big) \\
&-(\mu(t)(i))l_{i,j}(t) + (\mu(t)(i))\lambda_{i,j}(\mu(t))\Big)\Big]\ dt.
\end{aligned} \tag{6.7}
$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

From (6.6) and (6.5), we get $(\mu(t)(i))\ l_{i,j}(t) = T^{-1}(\nu(i) - \xi(i))g_{i,j}$, and this is nonzero only if $\nu(i) > \xi(i)$ and $\nu(j) < \xi(j)$; see (6.2). For convenience, let us define

$$\Upsilon = \{(i,j) \mid j \neq i,\ \nu(i) > \xi(i),\ \nu(j) < \xi(j)\}.$$

By assumptions (**A2**)-(**A3**), $|\log \lambda_{i,j}(\cdot)| \leq |\log C| + |\log c|$. Using these observations, (6.7) is upper bounded by

$$\begin{aligned}
(6.8)\quad & \int_{[0,T]} \Big[\sum_{(i,j)\in\Upsilon} \Big(T^{-1}(\nu(i) - \xi(i))g_{i,j} \log \Big(\frac{(\nu(i) - \xi(i))g_{i,j}}{T(\mu(t)(i))} \Big) \\
& \qquad + T^{-1}(\nu(i) - \xi(i))g_{i,j}(|\log C| + |\log c| + 1) + C \Big) \Big]\ dt \\
& \leq \sum_{(i,j)\in\Upsilon} (\nu(i) - \xi(i))g_{i,j} |\log((\nu(i) - \xi(i))g_{i,j})| \\
& \qquad - \sum_{i:\nu(i)>\xi(i)} (\nu(i) - \xi(i))T^{-1} \int_{[0,T]} \log(\mu(t)(i))\ dt \\
& \qquad + \quad ||\nu - \xi||_1 |\log T| \\
& \qquad + \quad ||\nu - \xi||_1 (|\log C| + |\log c| + 1) + CTr^2,
\end{aligned}$$

where in arriving at the last three terms we have repeatedly used (6.3). The quantity $||\nu - \xi||_1$ is the total variation distance between ξ and ν. Let us now bound the first two terms on the right-hand side of (6.8).

Observing that there is a constant K such that $\sup_{x\in[0,1]} x|\log x| \leq K < +\infty$, the first term on the right-hand side of (6.8) can be upper bounded as

$$\begin{aligned}
& \sum_{(i,j)\in\Upsilon} (\nu(i) - \xi(i))g_{i,j} |\log((\nu(i) - \xi(i))g_{i,j})| \\
& \leq \sum_{i:\nu(i)>\xi(i)} (\nu(i) - \xi(i))|\log(\nu(i) - \xi(i))| \Big(\sum_{j:(i,j)\in\Upsilon} g_{i,j} \Big) \\
& \qquad + \sum_{(i,j)\in\Upsilon} (\nu(i) - \xi(i))g_{i,j} |\log g_{i,j}| \\
& \leq \sum_{i:\nu(i)>\xi(i)} (\nu(i) - \xi(i))|\log(\nu(i) - \xi(i))| + K||\nu - \xi||_1 \\
(6.9)\quad & \leq \sum_i |(|\nu(i) - \xi(i)| \log |\nu(i) - \xi(i)|)| + K||\nu - \xi||_1.
\end{aligned}$$

To bound the second term on the right-hand side of (6.8), use (6.1) and employ the change of variable $u = \mu(t)(i)$ to get

$$
\begin{aligned}
-(\nu(i)-\xi(i))T^{-1}\int_{[0,T]}\log(\mu(t)(i))\ dt &= \int_{\nu(i)}^{\xi(i)}\log u\ du\\
&= [u\log u-u]_{\nu(i)}^{\xi(i)}\\
&\le |\xi(i)\log\xi(i)-\nu(i)\log\nu(i)|+|\nu(i)-\xi(i)|.
\end{aligned}
$$

Summing this over all i, we see that the second term in (6.8) is upper bounded by

$$
\sum_i |\xi(i)\log\xi(i)-\nu(i)\log\nu(i)|+||\nu-\xi||_1. \tag{6.10}
$$

Since $\mathcal{M}_1(\mathcal{Z})$ is compact, all terms in the upper bounds (6.9) and (6.10) are bounded. Substituting (6.9) and (6.10) on the right-hand side of (6.8), and noticing that $T>0$, we see that the right-hand side of (6.8) is upper bounded, and this upper bound serves as an upper bound on $S_{[0,T]}(\mu|\nu)$, which we summarize as

$$
\begin{aligned}
S_{[0,T]}(\mu|\nu) &\le \sum_i |(|\nu(i)-\xi(i)|\log|\nu(i)-\xi(i)|)|\\
&\quad+\sum_i |\xi(i)\log\xi(i)-\nu(i)\log\nu(i)|\\
&\quad+||\nu-\xi||_1(|\log T|)\\
&\quad+||\nu-\xi||_1(|\log C|+|\log c|+K+2)+CTr^2 \qquad (6.11)\\
&\le C_3(T)
\end{aligned}
$$

for a suitable constant $C_3(T)$ that is independent of ν and ξ. This concludes the proof of the first bullet for the case when all transitions are allowed.

When only those transitions in the directed edge set $\mathcal{E}$ can occur, since the Markov chain is irreducible (by assumption (**A1**)), there exists a finite sequence of intermediate points through which one can move from ν to ξ in $m=m(r,\mathcal{E})<+\infty$ steps:

$$
\nu=\nu^{(0)}\to\nu^{(1)}\to\cdots\to\nu^{(m)}=\xi.
$$

Consider now the piecewise linear path that moves from ν to ξ through the above sequence of points with velocities such that each segment is covered in time T/m. Then $S_{[0,T]}(\mu|\nu)\le C_1(T)=mC_3(T/m)$, and the proof of the first bullet is complete.

$S_T(\xi|\nu)\le C_1(T)$ follows immediately from (3.15), whence the second bullet follows.

To see the third bullet, we use (6.11). Since $\mathcal{M}_1(\mathcal{Z})$ is a subset of a finite dimensional space, the topology of weak convergence is the same as the

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

topology induced by the total variation metric. In particular, if $\rho_0(\nu,\xi) \to 0$ then $\nu(i) \to \xi(i)$ for every $i \in \mathcal{Z}$. As a consequence, for every $\varepsilon > 0$ and with $T = \varepsilon$, we can choose a $\delta > 0$ such that each of the first four terms in (6.11) is upper bounded by ε, and so $\rho_0(\nu,\xi) < \delta$ implies

$$S_\varepsilon(\xi|\nu) \le 4\varepsilon + Cr^2\varepsilon \le C_2\varepsilon$$

for some $C_2 < +\infty$, and the proof of the third bullet and the Lemma is complete. □

7. Proofs of Lemma 3.3 and Lemma 3.4. We begin with two useful lemmas.

LEMMA 7.1. *Let $L(t)$ be a matrix of rates such that the solution $\mu : [0,T] \to \mathcal{M}_1(\mathcal{Z})$ to the ODE $\dot\mu(t) = L(t)^*\mu(t)$ with $\mu(0) = \nu$ has $S_{[0,T]}(\mu|\nu) < +\infty$. There exists a constant $K < +\infty$ such that*

$$\int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ l_{i,j}(t) \Big]\ dt \le S_{[0,T]}(\mu|\nu) + KT.$$

PROOF. It is easy to verify that $\tau^*(u-1) = u\log u - u + 1 \ge u - e + 1$ for all $u \ge 0$. By Theorem 3.2, $S_{[0,T]}(\mu|\nu) < +\infty$ implies that its evaluation is given by (2.8). Using these two facts, we get

$$\begin{aligned}
S_{[0,T]}(\mu|\nu) &= \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ \lambda_{i,j}(\mu(t))\ \tau^*\left(\frac{l_{i,j}(t)}{\lambda_{i,j}(\mu(t))} - 1\right) \Big]\ dt \\
&\ge \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ \lambda_{i,j}(\mu(t)) \left(\frac{l_{i,j}(t)}{\lambda_{i,j}(\mu(t))} - e + 1\right) \Big]\ dt \\
&\ge \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ l_{i,j}(t) \Big]\ dt - (e-1)CrT,
\end{aligned}$$

and the lemma follows. □

The next lemma bounds the increase in the cost due to time scaling on a fixed path between two points.

LEMMA 7.2. *Let $L(t)$ be a matrix of rates such that the solution $\mu : [0,T] \to \mathcal{M}_1(\mathcal{Z})$ to the ODE $\dot\mu(t) = L(t)^*\mu(t)$ with $\mu(0) = \nu$ has $S_{[0,T]}(\mu|\nu) < +\infty$ and $\mu(T) = \xi$. Let $0 < \alpha < +\infty$ be a time scaling. With $T' = T/\alpha$, consider the path $\{\tilde\mu(t) = \mu(\alpha t) \mid t \in [0,T']\}$ having $\tilde\mu(0) = \nu$ and $\tilde\mu(T') = \xi$. Then*

$$\dot{\tilde\mu}(t) = \tilde L(t)^*\tilde\mu(t), \quad t \in [0,T']$$

where $\tilde{L}(t) = \alpha L(\alpha t)$. *Furthermore, the scaled path* $\tilde{\mu} : [0, T'] \to \mathcal{M}_1(\mathcal{Z})$ *satisfies*

$$S_{[0,T']}(\tilde{\mu}|\nu) \leq S_{[0,T]}(\mu|\nu) + |\log\alpha| \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i)) l_{i,j}(t) \Big]\, dt + \frac{|1-\alpha|}{\alpha} CrT. \tag{7.1}$$

Proof. Clearly $\tilde{\mu}(0) = \mu(0) = \nu$ and $\tilde{\mu}(T') = \mu(\alpha T') = \mu(T) = \xi$. Since $\dot{\mu}(t) = L(t)^*\mu(t)$, we also have

$$\dot{\tilde{\mu}}(t) = \frac{d\mu(\alpha t)}{dt} = \alpha\dot{\mu}(\alpha t) = \alpha L(\alpha t)^*\mu(\alpha t) = \alpha L(\alpha t)^*\tilde{\mu}(t)$$

from which $\tilde{L}(t) = \alpha L(\alpha t)$ is obvious. Its (i,j)th entry $\tilde{l}_{i,j}(t)$ equals $\alpha l_{i,j}(\alpha t)$. The cost of $\tilde{\mu} : [0, T'] \to \mathcal{M}_1(\mathcal{Z})$ is then

$$\begin{aligned}
S_{[0,T']}(\tilde{\mu}|\nu) &= \int_{[0,T']} \Big[\sum_{(i,j)\in\mathcal{E}} (\tilde{\mu}(t)(i))\ \lambda_{i,j}(\tilde{\mu}(t))\ \tau^* \left(\frac{\tilde{l}_{i,j}(t)}{\lambda_{i,j}(\tilde{\mu}(t))} - 1 \right) \Big] dt \\
&= \int_{[0,T']} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(\alpha t)(i))\ \lambda_{i,j}(\mu(\alpha t))\ \tau^* \left(\frac{\alpha l_{i,j}(\alpha t)}{\lambda_{i,j}(\mu(\alpha t))} - 1 \right) \Big] dt \\
&= \int_{[0,T']} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(\alpha t)(i))\ \lambda_{i,j}(\mu(\alpha t)) \\
&\quad \times\alpha \Big\{ \tau^* \left(\frac{l_{i,j}(\alpha t)}{\lambda_{i,j}(\mu(\alpha t))} - 1 \right) + \frac{l_{i,j}(\alpha t)}{\lambda_{i,j}(\mu(\alpha t))} (\log\alpha) + \frac{1-\alpha}{\alpha} \Big\} \Big] dt
\end{aligned}$$

where we have used the fact that

$$\tau^*(\alpha u - 1) = \alpha \left\{ \tau^*(u-1) + u(\log\alpha) + \frac{1-\alpha}{\alpha} \right\}, \quad u \geq 0.$$

Changing variables from αt to t and continuing, we get

$$\begin{aligned}
S_{[0,T']}(\tilde{\mu}|\nu) &= \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ \lambda_{i,j}(\mu(t))\ \tau^* \left(\frac{l_{i,j}(t)}{\lambda_{i,j}(\mu(t))} - 1 \right) \Big]\, dt \\
&\quad + (\log\alpha) \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ l_{i,j}(t) \Big]\, dt \\
&\quad + \frac{1-\alpha}{\alpha} \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\ \lambda_{i,j}(\mu(t)) \Big]\, dt.
\end{aligned}$$

Since the first term on the right-hand side above is $S_{[0,T]}(\mu|\nu)$ and $\lambda_{i,j}(\cdot) \leq C$, (7.1) follows. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

PROOF OF LEMMA 3.3. Fix $T > 0$. Fix an arbitrary ε such that $0 < \varepsilon < T/4$. Let $\delta > 0$ be as given by part 3 of Lemma 3.2 so that $\rho_0(\nu, \xi) < \delta$ implies $S_\varepsilon(\xi|\nu) \le C_2\varepsilon$.

Let $\{(\nu_i, \xi_i), i = 1, 2\}$ be two points in $\mathcal{Z} \times \mathcal{Z}$. By an abuse of notation, let ρ_T given by

$$\rho_T((\nu_1, \xi_1), (\nu_2, \xi_2)) = \max\{\rho_0(\nu_1, \nu_2), \rho_0(\xi_1, \xi_2)\}.$$

denote the metric on $\mathcal{Z} \times \mathcal{Z}$. Let $\rho_T((\nu_1, \xi_1), (\nu_2, \xi_2)) < \delta$. We need to show that $S_T(\xi_1|\nu_1)$ and $S_T(\xi_2|\nu_2)$ are close to each other.

Obviously, $\rho_0(\nu_1, \nu_2) < \delta$ and $\rho_0(\xi_2, \xi_1) < \delta$. Let μ denote the minimum cost path from ν_2 to ξ_2 in time T with cost $S_T(\xi_2|\nu_2)$. Consider the path from ν_1 to ξ_1 as follows:

- Traverse the path from ν_1 to ν_2 in time $[0, \varepsilon]$, as given by part 3 of Lemma 3.2. This traversal costs at most $C_2\varepsilon$.
- Given the optimal $[0, T]$-path μ from ν_2 to ξ_2, consider the sped-up path $\tilde{\mu} : [0, T - 2\varepsilon] \to \mathcal{M}_1(\mathcal{Z})$ given by $\tilde{\mu}(t) = \mu(\alpha t)$ with $\alpha = T/(T - 2\varepsilon)$. Travel from ν_2 to ξ_2 in the duration $[\varepsilon, T - \varepsilon]$ along the path $\tilde{\mu}$.
- Traverse the path from ξ_2 to ξ_1 in time $[0, \varepsilon]$, again as given by part 3 of Lemma 3.2. This traversal's cost is also at most $C_2\varepsilon$.

The minimum cost for traversal from ν_1 to ξ_1 is at most the sum of these paths. Hence, by Lemmas 7.2 and 7.1, we get

$$\begin{aligned} S_T(\xi_1|\nu_1) &\le 2C_2\varepsilon + S_T(\xi_2|\nu_2) + \Big(\log \frac{T}{T - 2\varepsilon}\Big)(S_T(\xi_2|\nu_2) + KT) + \frac{2\varepsilon}{T} CrT \\ &\le S_T(\xi_2|\nu_2) + 2C_2\varepsilon + \Big(\frac{T}{T - 2\varepsilon} - 1\Big)(C_1(T) + KT) + 2Cr\varepsilon \end{aligned}$$

where we used $\log u \le u - 1$ for $u > 0$. Observing that

$$\varepsilon < T/4 \Rightarrow \frac{T}{T - 2\varepsilon} - 1 = \frac{2\varepsilon}{T - 2\varepsilon} \le \frac{4\varepsilon}{T},$$

we deduce that

$$S_T(\xi_1|\nu_1) \le S_T(\xi_2|\nu_2) + C_4(T)\varepsilon$$

where we may take $C_4(T) = 2C_2 + 4K + 4C_1(T)/T + 2Cr$. Reversing the roles of (ν_1, ξ_1) and (ν_2, ξ_2), we deduce

$$|S_T(\xi_1|\nu_1) - S_T(\xi_2|\nu_2)| \le C_4(T)\varepsilon.$$

This concludes the proof that $(\nu, \xi) \mapsto S_T(\xi|\nu)$ is uniformly continuous. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

We now provide the proof of the result on uniform continuity of the analogous quantity $V(\xi|\nu)$.

PROOF OF LEMMA 3.4. Fix $\varepsilon > 0$ and choose δ as in the third part of Lemma 3.2. Let (ν_1, ξ_1) and (ν_2, ξ_2) be such that the starting points are δ-close to each other and so are the ending points, that is, $\rho_0(\nu_1, \nu_2) < \delta$ and $\rho_0(\xi_1, \xi_2) < \delta$. Consider the following path:

- Traverse the path from ν_1 to ν_2 in time $[0, \varepsilon]$, as given by part 3 of Lemma 3.2. This traversal costs at most $C_2\varepsilon$.
- Traverse by a path from ν_2 to ξ_2 in finite time by a path with cost at most $V(\xi_2|\nu_2) + C_2\varepsilon$.
- Traverse the path from ξ_2 to ξ_1 in time $[0, \varepsilon]$, again as given by part 3 of Lemma 3.2. This traversal's cost is also at most $C_2\varepsilon$.

We then have

$$V(\xi_1|\nu_1) \leq C_2\varepsilon + (V(\xi_2|\nu_2) + C_2\varepsilon) + C_2\varepsilon = V(\xi_2|\nu_2) + 3C_2\varepsilon.$$

Reversing the roles of (ν_1, ξ_1) and (ν_2, ξ_2) and via a similar argument, we deduce that

$$|V(\xi_1|\nu_1) - V(\xi_2|\nu_2)| \leq 3C_2\varepsilon$$

which shows that $(\nu, \xi) \mapsto V(\xi|\nu)$ is uniformly continuous. □

8. Proof of Theorem 3.4. Again, we proceed through a sequence of lemmas. Let $\nu_N \to \nu$ weakly. By Theorem 3.2, the sequence of laws of the terminal measure $(p^{(N)}_{\nu_N,T}, N \geq 1)$ satisfies the large deviation principle with speed N and good rate function $S_T(\xi|\nu)$. By Varadhan's lemma, for every $f \in C_b(\mathcal{M}_1(\mathcal{Z}))$, we have

$$\lim_{N\to+\infty} \frac{1}{N} \log \int_{\mathcal{M}_1(\mathcal{Z})} e^{Nf}\, dp^{(N)}_{\nu_N,T} = \sup_{\xi\in\mathcal{M}_1(\mathcal{Z})} [f(\xi) - S_T(\xi|\nu)]. \tag{8.1}$$

Let us define

$$\Lambda(f|\nu) = \sup_{\xi\in\mathcal{M}_1(\mathcal{Z})} [f(\xi) - S_T(\xi|\nu)]. \tag{8.2}$$

Observe that the rate function admits the characterization (see, e.g., [13, Th. 4.4.2])

$$S_T(\xi|\nu) = \sup_{f\in C_b(\mathcal{M}_1(\mathcal{Z}))} [f(\xi) - \Lambda(f|\nu)]. \tag{8.3}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA 8.1. *Let $f \in C_b(\mathcal{M}_1(\mathcal{Z}))$. The mapping $\nu \in \mathcal{M}_1(\mathcal{Z}) \mapsto \Lambda(f|\nu) \in \mathbb{R}$ is continuous.*

PROOF. Since f is continuous, by Lemma 3.3, the mapping

$$\eta : (\nu, \xi) \in \mathcal{M}_1(\mathcal{Z}) \times \mathcal{M}_1(\mathcal{Z}) \mapsto f(\xi) - S_T(\xi|\nu) \in \mathbb{R}$$

is jointly continuous. As $\mathcal{M}_1(\mathcal{Z})$ is compact, the supremum in the definition of (8.2) is attained.

Let $\nu_N \to \nu$ weakly, and for each ν_N, let ξ_N denote a point where the supremum in the definition of (8.2) is attained. In other words, $\Lambda(f|\nu_N) = \eta(\nu_N, \xi_N)$ for each N. The sequence $((\nu_N, \xi_N), N \geq 1)$ has a convergent subsequence that converges to (ν, ξ), for some ξ. Reindex so that we may take $(\nu_N, \xi_N) \to (\nu, \xi)$ as $N \to +\infty$. By the continuity of η,

$$\Lambda(f|\nu_N) = \eta(\nu_N, \xi_N) \to \eta(\nu, \xi)$$

as $N \to +\infty$. The proof will be complete if we can show that $\Lambda(f|\nu) = \eta(\nu, \xi)$, that is, the supremum in $\eta(\nu, \cdot)$ is attained at ξ.

To see this, observe that for any ξ', we have $\eta(\nu_N, \xi') \leq \eta(\nu_N, \xi_N)$, and so

$$\eta(\nu, \xi') = \lim_{N \to +\infty} \eta(\nu_N, \xi') \leq \limsup_{N \to +\infty} \eta(\nu_N, \xi_N) = \eta(\nu, \xi).$$

This completes the proof of the lemma. □

Our next result show that the convergence in (8.1) is uniform. This is where our uniform large deviation result for nonchaotic initial conditions comes in handy.

Recall that $\mathcal{M}_1^{(N)}(\mathcal{Z}) \subset \mathcal{M}_1(\mathcal{Z})$ is the subset of values taken by the initial empirical measure ν when there are N particles.

LEMMA 8.2. *The convergence in (8.1) is uniform in the following sense: for each $f \in C_b(\mathcal{M}_1(\mathcal{Z}))$, we have*

$$\lim_{N \to +\infty} \sup_{\nu \in \mathcal{M}_1^{(N)}(\mathcal{Z})} \left| \frac{1}{N} \log \int_{\mathcal{M}_1(\mathcal{Z})} e^{Nf} \, dp_{\nu,T}^{(N)} - \Lambda(f|\nu) \right| = 0. \tag{8.4}$$

PROOF. We will prove this by contradiction. Suppose the above limit is not zero. Then there is an $\varepsilon > 0$ and an infinite subset $\mathbb{V}_1 \subset \mathbb{N}$ such that

$$\sup_{\nu \in \mathcal{M}_1^{(N)}(\mathcal{Z})} \left| \frac{1}{N} \log \int_{\mathcal{M}_1(\mathcal{Z})} e^{Nf} \, dp_{\nu,T}^{(N)} - \Lambda(f|\nu) \right| > \varepsilon, \quad \text{for every } N \in \mathbb{V}_1,$$

that is, the violations occur infinitely often. So we can find a sequence $(\nu_N)_{N\in\mathbb{V}_1}$ such that

$$\left|\frac{1}{N}\log\int_{\mathcal{M}_1(\mathcal{Z})} e^{Nf}\ dp^{(N)}_{\nu_N,T} - \Lambda(f|\nu_N)\right| > \varepsilon, \quad \text{for every } N \in \mathbb{V}_1.$$

Extract a further subsequence, which is another infinite subset $\mathbb{V}_2 \subset \mathbb{V}_1$, such that $(\nu_N)_{N\in\mathbb{V}_2} \to \nu$ for some ν. By Lemma 8.1, $(\Lambda(f|\nu_N))_{N\in\mathbb{V}_2} \to \Lambda(f|\nu)$, and so

(8.5)
$$\left|\frac{1}{N}\log\int_{\mathcal{M}_1(\mathcal{Z})} e^{Nf}\ dp^{(N)}_{\nu_N,T} - \Lambda(f|\nu)\right| > \frac{\varepsilon}{2}, \text{ for all sufficiently large } N \in \mathbb{V}_2.$$

Construct a new initial state sequence $(\nu_N)_{N\geq 1}$ that matches with the above subsequence for $N \in \mathbb{V}_2$ and such that $\nu_N \to \nu$. For such a sequence, by Theorem 3.2 and Varadhan's lemma, (8.1) holds. But (8.5) for all sufficiently large $N \in \mathbb{V}_2$ is a contradiction to (8.1). □

Proof of Theorem 3.4. Consider the joint measure $\wp^{(N)}_{0,T}$ given by

$$d\wp^{(N)}_{0,T}(\nu,\xi) = d\wp^{(N)}_0(\nu)dp^{(N)}_{\nu,T}(\xi).$$

We shall apply Feng and Kurtz's [18, Prop. 3.25]. To do this, we need to verify three conditions listed below.

- *Exponential tightness.* The sequence $(\wp^{(N)}_{0,T}, N \geq 1)$, which comprises of probability measures on the compact product space, is trivially exponentially tight.
- *Uniform convergence of the Laplace-Varadhan functional in the initial condition.* For each N, the probability measure $p^{(N)}_{\nu,T}$ is supported on the compact subset $\mathcal{M}^{(N)}_1(\mathcal{Z})$. By Lemma 8.2, for each $f \in C_b(\mathcal{M}_1(\mathcal{Z}))$, the convergence of the Laplace-Varadhan functional is uniform, as given in (8.4).
- *Continuity of Laplace-Varadhan functional in the initial condition.* For each $f \in C_b(\mathcal{M}_1(\mathcal{Z}))$, the function $\nu \mapsto \Lambda(f|\nu)$ is continuous, by Lemma 8.1.

Under the above conditions, Feng and Kurtz demonstrated in [18, Prop. 3.25 and Rem. 3.26] that if the first marginal sequence $(\wp^{(N)}_0, N \geq 1)$ satisfies the large deviation principle with speed N and good rate function s, then so does the sequence of joint laws $(\wp^{(N)}_{0,T}, N \geq 1)$ with good rate function $S_{0,T}(\nu,\xi) = s(\nu) + S_T(\xi|\nu)$. This concludes the proof. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

APPENDIX A: MULTIPLE ω-LIMIT SETS

In this appendix, we verify that the classical program of Freidlin-Wentzell [21, Ch. 6] can be extended to our setting. There are primarily two things to keep in mind. First, for all finite N, we have a jump process on the simplex. Second, the quantity $V(\cdot|\cdot)$ defined in (2.9) is only uniformly continuous and not Lipschitz continuous. But this uniform continuity suffices. Though the changes are minor, we provide the entire sequence of lemmas with modified proofs for completeness and ease of verification.

A.1. Auxiliary results. We begin with a subset of the auxiliary results in [21, Ch. 6] that were shown for diffusions on a compact manifold.

LEMMA A.1. (Freidlin and Wentzell [21, Ch.6, Lemma 1.2]) *For any $\varepsilon > 0$ and any compact set $K \subset \mathcal{M}_1(\mathcal{Z})$, there exists a T_0 such that for any $\nu, \xi \in K$ there exists a function $\mu(t), t \in [0,T], \mu(0) = \nu, \mu(T) = \xi, T \le T_0$ with $S_{[0,T]}(\mu|\nu) \le V(\xi|\nu) + \varepsilon$.*

PROOF. Fix $\varepsilon > 0$. By part 3 of Lemma 3.2, with the constant C_2 as in that Lemma, and $\varepsilon_1 = \varepsilon/(4C_2)$, there is a $\delta_1 \in (0, \varepsilon_1)$ such that two points within a distance δ_1 can be connected by a path of duration ε_1 and cost at most $C_2\varepsilon_1 = \varepsilon/4$.

By Lemma 3.4, there is a δ_2 such that

$$\max\{\rho_0(\xi_1,\xi_2), \rho_0(\nu_1,\nu_2)\} < \delta_2 \text{ implies } |V(\xi_2|\nu_2) - V(\xi_1|\nu_1)| \le \varepsilon/4.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Choose a finite δ-net $\{\nu_i\}$ of points in K. Connect them with curves $\mu_{i,j}(t), t \in [0, T_{i,j}], \mu_{i,j}(0) = \nu_i, \mu_{i,j}(T_{i,j}) = \nu_j$ such that

$$S_{[0,T_{i,j}]}(\mu_{i,j}|\nu_i) \le V(\nu_j|\nu_i) + \varepsilon/4.$$

For arbitrary $\nu, \xi \in K$, let ν_k and ν_l be the respective closest points on the net. We can now find a path from ν to ν_k, then to ν_l along $\mu_{k,l}$, and then to ξ, with overall cost at most

$$\varepsilon/4 + (V(\nu_l|\nu_k) + \varepsilon/4) + \varepsilon/4 = V(\nu_l|\nu_k) + 3\varepsilon/4 \le V(\xi|\nu) + \varepsilon$$

where the last inequality follows from the uniform continuity in Lemma 3.4 and the choice of δ. The duration of the path is $T_{k,l} + 2\varepsilon_1 \le T_0 := (\max_{i,j} T_{i,j}) + 2\varepsilon_1$. □

For a set $A \subset \mathcal{M}_1(\mathcal{Z})$, let $[A]_\delta$ denote the (open) δ-neighborhood of A. Its closure will be denoted $\overline{[A]_\delta}$. For the following lemma, recall the notion of equivalence between two points: $\nu \sim \xi$ if $V(\xi|\nu) = V(\nu|\xi) = 0$ (see Section 4.2).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA A.2. (Freidlin and Wentzell [21, Ch.6, Lemma 1.6]) *Let all points of a compact set K be equivalent to each other, but not to any other point in $\mathcal{M}_1(\mathcal{Z})$. For any $\varepsilon > 0$, $\delta > 0$, $\nu, \xi \in K$, there exists a $T > 0$ and a function $\mu(t), 0 \le t \le T$ with $\mu(0) = \nu, \mu(T) = \xi, \mu(t) \in [K]_\delta$ for all $t \in [0,T]$, and $S_{[0,T]}(\mu|\nu) < \varepsilon$.*

PROOF. Fix $\varepsilon > 0, \delta > 0, \nu, \xi \in K$. We can find a sequence $(T_n, n \ge 1)$ and paths $\mu^{(n)} : [0, T_n] \to \mathcal{M}_1(\mathcal{Z})$ such that $\mu^{(n)}(0) = \nu, \mu^{(n)}(T_n) = \xi$ for all $n \ge 1$, and $\varepsilon > S_{[0,T_n]}(\mu^{(n)}|\nu) \to 0$ as $n \to +\infty$. Observe that $\mathcal{M}_1(\mathcal{Z}) \setminus [K]_\delta$ is compact, and so if an infinite number of $\mu^{(n)}$ left $[K]_\delta$, there is a limit point z outside $[K]_\delta$. Using part 3 of Lemma 3.2, and $S_{[0,T_n]}(\mu^{(n)}|\nu) \to 0$ as $n \to +\infty$, it follows that $V(z|\nu) = V(\xi|z) = 0$. Together with $V(\nu|\xi) = 0$, we conclude that $V(\nu|z) = 0$ and so $z \sim \nu$. But then z is an equivalent point outside K, which is a contradiction.

Hence $\mu^{(n)}$ goes outside $[K]_\delta$ for finitely many n. Take the first index larger than these. The corresponding path remains completely within $[K]_\delta$, and meets all the other requirements. □

In this section, we shall use the notation

$$\tau_A := \inf\{t \ge 0 \mid \mu_N(t) \notin A\}.$$

The law for this exit time depends on N and ν through the law $p_\nu^{(N)}$ for μ_N. This dependence will be assumed as understood and will be suppressed for brevity.

LEMMA A.3. (Freidlin and Wentzell [21, Ch.6, Lemma 1.7]). *Let all points of a compact set K be equivalent to each other and let $K \ne \mathcal{M}_1(\mathcal{Z})$. For a $\delta > 0$, let*

$$\tau_{[K]_\delta} := \inf\{t \ge 0 \mid \mu_N(t) \notin [K]_\delta\}.$$

For any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all sufficiently large N and all $\nu \in [K]_\delta$, we have

$$\mathbb{E}[\tau_{[K]_\delta}] < e^{+N\varepsilon}$$

where the expectation is with respect to the measure $p_\nu^{(N)}$.

PROOF. Fix $\varepsilon > 0$. Again by part 3 of Lemma 3.2, there is a $\delta_1 > 0$ such that two points δ_1-close have a path connecting them of duration $\varepsilon_1 = \varepsilon/(4C_2)$ and cost at most $\varepsilon/4$. Choose ξ outside K such that $\rho_0(\xi, K) < \delta_1$. Choose $\delta < \rho_0(\xi, K)/2$; we thus have $0 < \delta < \rho_0(\xi, K)/2 < \rho_0(\xi, K) < \delta_1$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Consider a finite δ-net of K. Lemma A.1 assures existence of paths that connect any pair of the net with cost at most $\varepsilon/4$. Let T_0' denote the maximum time among these paths, where the maximum is over pairs belonging to the net, and let $T_0 = T_0' + 2\varepsilon_1$. Traverse from any $\nu \in [K]_\delta$ to its nearest point on the net, then traverse from that point to the point on the net nearest to ξ, and thence to ξ. Now extend this path following the McKean-Vlasov dynamics so that the total duration is now T_0. This last appendage incurs no additional cost. Denote by μ the resulting path of duration T_0. Clearly $S_{[0,T_0]}(\mu|\nu) \leq 3\varepsilon/4$.

Now, any trajectory that is strictly δ-close to the trajectory μ exits $[K]_\delta$ at least once in the interval $[0, T_0]$ because for some $t \in [0, T_0]$, we have $\mu(t) = \xi$ which is at a distance greater than 2δ from K. We then have

$$\begin{aligned} p_\nu^{(N)}\left\{\tau_{[K]_\delta} < T_0\right\} &\geq p_\nu^{(N)}\left\{\rho_{T_0}(\mu_N, \mu) < \delta\right\} \\ &\geq e^{-3\varepsilon N/4}, \text{ for all } \nu \in [K]_\delta, \text{ for all } N \geq \text{ some } N_0, \end{aligned}$$

where the last inequality holds by (3.14) in Corollary 3.1. Consequently

$$p_\nu^{(N)}\left\{\tau_{[K]_\delta} \geq T_0\right\} \leq 1 - e^{-3\varepsilon N/4}, \quad \text{for all } \nu \in [K]_\delta, N \geq N_0.$$

This uniform bound, the Markov property, and induction imply

$$p_\nu^{(N)}\left\{\tau_{[K]_\delta} \geq mT_0\right\} \leq \left(1 - e^{-3\varepsilon N/4}\right)^m,$$

from which we obtain

$$\mathbb{E}[\tau_{[K]_\delta}] \leq T_0 \sum_{m\geq 0} (1 - e^{-3\varepsilon N/4})^m = T_0 e^{3\varepsilon N/4} < e^{\varepsilon N}$$

where the last inequality holds for all sufficiently large N. This concludes the proof. □

LEMMA A.4. (Freidlin and Wentzell [21, Ch.6, Lemma 1.8]). *Let K be an arbitrary compact subset of $\mathcal{M}_1(\mathcal{Z})$ and let G be a neighborhood of K. For any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all sufficiently large N and all ν belonging to $\overline{g}$, with $g = [K]_\delta$ and $\overline{g} = \overline{[K]_\delta}$, we have*

$$\mathbb{E}\left[\int_{[0,\tau_G]} \mathbf{1}_g\left(\mu_N(t)\right)\, dt\right] > e^{-\varepsilon N},$$

where the expectation is with respect to the measure $p_\nu^{(N)}$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

PROOF. Fix $\varepsilon > 0$. Choose δ_1 small enough so that $[K]_{\delta_1} \subset G$. Next, choose δ_2 as in part 3 of Lemma 3.2 so that with $\varepsilon_1 = \varepsilon/(2C_2)$, any two δ_2-close points can be connected by a path of duration ε_1 and cost at most $\varepsilon/2$. Now let $\delta < \min\{\delta_1, \delta_2/2\}$ and set $g = [K]_\delta$.

Fix $T > \varepsilon_1$. Take any $\nu \in \overline{g}$. Connect it to the closest point on K (via a path of duration ε_1 and cost $\leq \varepsilon/2$) and then extend via the McKean-Vlasov path with this initial condition for a further duration of $T - \varepsilon_1$. Call the entire path of duration T as μ. So long as μ is inside the $\delta/3$ neighborhood of K, any $\delta/3$ neighborhood of μ lies completely inside $g = [K]_\delta$. By assumptions (**A2**)-(**A3**), the McKean-Vlasov dynamics has a bounded velocity field. Consequently, the part of μ that begins at K and until either its exit from $\overline{[K]_{\delta/3}}$ or time T, whichever occurs earlier, is of duration at least t_0 for some $t_0 > 0$, independent of the starting point. It follows that

$$\{\rho_T(\mu_N, \mu) \leq \delta/3\} \text{ implies } \{\tau_g \geq \min\{T, t_0\}\}.$$

Furthermore, $\tau_G \geq \tau_g$ and $\mathbf{1}_g(\mu_N(t)) = 1$ until the random path exits g. Thus

$$\begin{aligned}
\mathbb{E}\left[\int_{[0,\tau_G]} \mathbf{1}_g(\mu_N(t))\, dt\right] &\geq \mathbb{E}[\tau_g] \\
&\geq \mathbb{E}[\tau_g \cdot \mathbf{1}\{\rho_T(\mu_N, \mu) \leq \delta/3\}] \\
&\geq \min\{T, t_0\} \cdot p_\nu^{(N)}\{\rho_T(\mu_N, \mu) \leq \delta/3\} \\
&\geq \min\{T, t_0\} \cdot e^{-N\varepsilon/2} \quad \text{(by (3.14))} \\
&\geq e^{-N\varepsilon},
\end{aligned}$$

where the last two inequalities hold for all sufficiently large N. □

LEMMA A.5. (Freidlin and Wentzell [21, Ch.6, Lemma 1.9]). *Let K be a compact subset of $\mathcal{M}_1(\mathcal{Z})$ not containing any ω-limit set entirely. There exist positive constants c and T_0 such that for all sufficiently large N, any $T > T_0$, and any $\nu \in K$, we have*

$$p_\nu^{(N)}\{\tau_K > T\} \leq e^{-Nc(T-T_0)}.$$

PROOF. For a sufficiently small δ, the closed δ-neighborhood of K, denoted $\overline{[K]_\delta}$, does not contain any ω-limit set entirely. Indeed, if this were not true, we can find a sequence of $\delta \downarrow 0$ such that each set in the nested decreasing sequence of sets $\overline{[K]_\delta}$ contains an ω-limit set entirely. For a δ, define $\Omega(\delta)$ to be the closure of the union of ω-limit sets contained in $\overline{[K]_\delta}$. Clearly, $\Omega(\delta)$ is a positively invariant set, and the family indexed by δ is

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

a nested decreasing sequence of nonempty compact sets. Then $\cap_\delta \Omega(\delta)$ is a nonempty compact invariant set in K. Further, it contains an ω-limit set entirely, a contradiction.

For $\nu \in \overline{[K]_\delta}$, denote by $\tau(\nu)$ the time for first exit of the solution to the McKean-Vlasov equation with initial condition ν from the set $\overline{[K]_\delta}$. Since $\overline{[K]_\delta}$ does not contain any ω-limit set entirely, $\tau(\nu) < +\infty$ for all $\nu \in \overline{[K]_\delta}$. The function $\tau(\nu)$ is upper semicontinuous, and consequently, it attains its largest value $\max_{\nu \in \overline{[K]_\delta}} \tau(\nu) = T_1 < +\infty$.

Set $T_0 = T_1 + 1$ and consider all paths of duration T_0 that take values only in $\overline{[K]_\delta}$. It is easy to see that this set is closed. It follows that for each $\nu \in \overline{[K]_\delta}$, we have that $S_{[0,T_0]}(\cdot|\nu)$ attains its minimum $A(\nu)$ on this set. Further, the mapping $\nu \mapsto A(\nu)$ is continuous, as can be shown by an easy application of Lemmas 3.2, 7.1 and 7.2. Thus $A := \min_{\nu \in \overline{[K]_\delta}} A(\nu)$ is attained. This minimum is strictly positive since there are no trajectories of the McKean-Vlasov equation among the paths under consideration.

Fix $\varepsilon < A$, a $\nu \in K$, and consider the family of paths

$$\Phi_\nu(A - \varepsilon/2) = \{\mu : [0, T_0] \to \mathcal{M}_1(\mathcal{Z}) \mid S_{[0,T_0]}(\mu|\nu) \le A - \varepsilon/2\}.$$

Any path in this set exits $\overline{[K]_\delta}$ in the interval $[0, T_0]$. With initial state ν, the event $\{\tau_K > T_0\}$ implies that the trajectory remains entirely within K, and since any path in $\Phi_\nu(A - \varepsilon/2)$ exits $\overline{[K]_\delta}$, we must have

$$\rho_{T_0}(\mu_N, \Phi_\nu(A - \varepsilon/2)) \ge \delta.$$

It follows that

$$p_\nu^{(N)}\{\tau_K > T_0\} \quad \le \quad p_\nu^{(N)}\{\rho_{T_0}(\mu_N, \Phi_\nu(A - \varepsilon/2)) \ge \delta\}.$$

By considering any $\nu_1 \in K$, and by using (3.13) of Corollary 3.1, we have

$$\begin{aligned}
\sup_{\nu_1 \in K} p_{\nu_1}^{(N)}\{\tau_K > T_0\} &\le \sup_{\nu_1 \in K} p_{\nu_1}^{(N)}\{\rho_{T_0}(\mu_N, \Phi_\nu(A - \varepsilon/2)) \ge \delta\} \\
&\le \exp\{-N(\inf\{S_{[0,T]}(\mu|\nu_1) | \mu \in \Phi_{\nu_1}(A - \varepsilon/2)\} - \varepsilon/2)\} \\
&\qquad \text{for all sufficiently large } N \\
&\le e^{-N(A-\varepsilon)} \qquad \text{for all sufficiently large } N.
\end{aligned}$$

For our fixed $\nu \in K$, the Markov property then implies

$$\begin{aligned}
p_\nu^{(N)}\{\tau_K > (m+1)T_0\} &\le \mathbb{E}[\mathbf{1}\{\tau_K > mT_0\} \cdot \mathbb{E}[\mathbf{1}\{\tau_K > T_0\} \mid \mu^{(N)}(mT_0)]] \\
&\le p_\nu^{(N)}\{\tau_K > mT_0\} \cdot \left(\sup_{\nu_1 \in K} p_{\nu_1}^{(N)}\{\tau_K > T_0\} \right) \\
&\le p_\nu^{(N)}\{\tau_K > mT_0\} \cdot e^{-N(A-\varepsilon)}.
\end{aligned}$$

By induction, for a $T > T_0$, we have

$$\begin{aligned} p_\nu^{(N)}\{\tau_K > T\} &\leq p_\nu^{(N)}\left\{\tau_K > \left\lfloor \frac{T}{T_0} \right\rfloor T_0\right\} \\ &\leq e^{-N(A-\varepsilon)\left(\left\lfloor \frac{T}{T_0} \right\rfloor\right)} \\ &\leq e^{-N(A-\varepsilon)(T/T_0-1)} = e^{-Nc(T-T_0)} \end{aligned}$$

for $c = (A-\varepsilon)/T_0$, and this completes the proof. □

The above theorem has the following immediate corollary.

COROLLARY A.1. (Freidlin and Wentzell [21, Ch.6, Corollary to Lemma 1.9]). *Let K be a compact set not containing any ω-limit set entirely. There exists a positive integer N_0 and a positive constant c such that for $N \geq N_0$ and any $\nu \in K$, we have*

$$\mathbb{E}[\tau_K] \leq T_0 + 1/(cN_0)$$

where the expectation is with respect to the measure $p_\nu^{(N)}$.

Recall the definition of V given in (2.9), and the notion of equivalence on $\mathcal{M}_1(\mathcal{Z})$ given in Section 4.2. Under condition (**B**) in Section 4.2, we have equivalent sets $K_1, \ldots, K_l$ to which all ω-limit sets converge. We shall now define a discrete-time Markov chain of states at hitting times of neighborhoods of these compact sets. In order to bound the transition probabilities of this chain, recall the definitions of $\tilde{V}(K_i, K_j)$ given in (2.10) and $V(K_i, K_j)$ given in (2.11).

Define the following quantities:

- An r_0 such that $0 < r_0 < (1/2)\min_{i,j} \rho_0(K_i, K_j)$,
- An r_1 such that $0 < r_1 < r_0$,
- The set C as $C := \overline{\mathcal{M}_1(\mathcal{Z})} \setminus \cup_{i=1}^l [K_i]_{r_0}$,
- The set Γ_i as $\Gamma_i := \overline{[K_i]_{r_0}}$,
- The set g_i as $g_i := [K_i]_{r_1}$, and finally,
- The set g as $g := \cup_{i=1}^l g_i$.

Let us now define the following stopping times:

- $\tau_0 := 0$,
- The time for exit from the union of the r_0 neighborhoods of the compact sets K_i's, that is, $\sigma_n := \inf\{t \geq \tau_n \mid \mu_N(t) \in C\}$,
- The time to re-enter $\overline{g}$, that is, $\tau_n := \inf\{t \geq \sigma_n \mid \mu_N(t) \in \overline{g}\}$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Finally, we define $Z_n := \mu_N(\tau_n)$. We shall use the notation $p^{(N)}(\nu, \overline{g_j})$ for $p^{(N)}_\nu(\mu_N(\tau_n) \in \overline{g_j})$ when $\mu_N(\tau_{n-1}) = \nu$.

LEMMA A.6. (Freidlin and Wentzell [21, Ch.6, Lemma 2.1]). *For any $\varepsilon > 0$, there is a small enough $r_0 > 0$ such that for any r_2 satisfying $0 < r_2 < r_0$, there is an r_1 satisfying $0 < r_1 < r_2$ such that for all sufficiently large N, for all $\nu \in \overline{[K_i]_{r_2}}$, the one-step transition probabilities of Z_n satisfy*

$$\exp\{-N(\tilde{V}(K_i, K_j) + \varepsilon)\} \le p^{(N)}(\nu, \overline{g_j}) \le \exp\{-N(\tilde{V}(K_i, K_j) - \varepsilon)\}.$$

PROOF. For pairs with $\tilde{V}(K_i, K_j) = +\infty$, there is no smooth curve from K_i to K_j without touching one of the other compact sets. It follows that for any arbitrary $0 < r_1 < r_2 < r_0$, for all sufficiently large N, there is no path in $\mathcal{M}_1^{(N)}(\mathcal{Z})$ from $[K_i]_{r_2}$ to $\overline{g}_j$ without touching $[K_{i'}]_{r_0}$ for some $i' \ne i, j$. Thus, for all sufficiently large N, we have $p^{(N)}(\nu, \overline{g}_j) = 0$. The validity of the lemma is obvious for such pairs.

For all other pairs $\tilde{V}(K_i, K_j) \le V_0$ for some $V_0 < +\infty$.

Let us first argue the lower bound. Fix $\varepsilon > 0$. Choose δ as in part 3 of Lemma 3.2 so that with $\varepsilon_1 = \varepsilon/(10C_2)$, any two points δ-close can be connected by a path of duration ε_1 and cost at most $\varepsilon/10$. Set

$$r_0 = \min\{\delta/2, (1/3) \min_{i,j} \rho_0(K_i, K_j)\}.$$

Fix arbitrary r_2 satisfying $0 < r_2 < r_0$.

For each (i, j) with $\tilde{V}(K_i, K_j) < +\infty$, choose paths $\mu^{i,j} : [0, T_{i,j}] \to \mathcal{M}_1(\mathcal{Z})$ such that

- $\mu^{i,j}(0) \in K_i$,
- $\mu^{i,j}(T_{i,j}) \in K_j$,
- $\mu^{i,j}(t)$ does not touch $\cup_{i' \ne i,j} K_{i'}$ for $t \in [0, T_{i,j}]$, and
- $S_{[0,T]}(\mu^{i,j} | \mu^{i,j}(0)) \le \tilde{V}(K_i, K_j) + 0.2\varepsilon$.

Now choose r_1 so that

$$r_1 < \min\left\{r_2, \frac{r_0}{2}, \frac{1}{2} \min\left\{\rho_0\left(\mu^{i,j}(t), \cup_{i' \ne i,j} K_{i'}\right) \mid t \in [0, T_{i,j})], 1 \le i, j \le l\right\}\right\}.$$

Also choose $\delta' < \min\{r_0 - r_2, r_1\}$ so that $r_2 + \delta' < r_0$ and hence $r_0 + \delta' < 2r_0 < \delta$ by the choice of δ.

Take any $\nu \in \overline{[K_i]_{r_2}}$. Fix a finite δ'-net of K_i. If $i \ne j$, consider the following path.

- Connect ν to the nearest point $\nu_1 \in K_i$ with a path of duration ε_1 and cost at most 0.1ε.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

- Connect ν_1 to the nearest point ν_2 on the δ'-net of K_i again with a path of duration ε_1 and cost at most 0.1ε.
- Let ν_3 be the point on the δ'-net nearest to $\mu^{i,j}(0)$. Traverse the path given by Lemma A.2 that connects ν_2 to ν_3 without leaving the r_1-neighborhood of K_i. Thanks to the finite number of points on the δ'-net, this can be done in bounded time. Moreover, the cost is at most 0.1ε.
- Connect ν_3 to $\mu^{i,j}(0)$ with path of duration ε_1 and cost at most 0.1ε.
- Then traverse the path given by $\mu^{i,j}$.

If $i = j$, then simply take ν to a point at a distance $r_0 + \delta'$ from K_i and then to the nearest point in K_i. Note that $r_0 + \delta' < \delta$, and so the duration of this path is $2\varepsilon_1$ and cost at most 0.2ε. The constructed path is of bounded time duration, bounded say by T_0. We can thus extend all paths to duration T_0 along the McKean-Vlasov path, an appendage that incurs no additional cost. Call the resulting path $\mu(t), t \in [0, T_0]$. Clearly,

$$S_{[0,T_0]}(\mu|\nu) \leq \tilde{V}(K_i, K_j) + 0.6\varepsilon.$$

If $\rho_{T_0}(\mu_N, \mu) < \delta'$, then the trajectory μ_N begins at a point ν within an r_0-neighborhood of K_i, reaches the δ'-neighborhood of K_j and so hits $\overline{g}_j = \overline{[K_j]_{r_1}}$, is at most δ' distance away from the trajectory μ, and hence does not hit the r_0-neighborhood of any $K_{i'}, i' \neq i, j$; then $\mu_N(\tau_n) \in \overline{g}_j$. In other words,

$$\{\rho_{T_0}(\mu_N, \mu) < \delta'\} \subset \{\mu_N(\tau_n) \in \overline{g}_j\},$$

and so

$$\begin{aligned} p^{(N)}(\nu, \overline{g}_j) &\geq p_\nu^{(N)}\{\rho_{T_0}(\mu_N, \mu) < \delta'\} \\ &\geq \exp\{-N(S_{[0,T_0]}(\mu|\nu) + 0.1\varepsilon)\} \\ &\geq \exp\{-N(\tilde{V}(K_i, K_j) + \varepsilon)\}, \end{aligned}$$

where the second inequality holds for all sufficiently large N uniformly over the initial condition, thanks to (3.14) of Corollary 3.1. This establishes the lower bound.

We now prove the upper bound. Consider any path μ of some duration T starting at ν in the r_1-neighborhood of K_i, ending at a point say ξ in the δ'-neighborhood of $\overline{g}_j$ at time T, and not touching any of the other compact sets $K_{i'}, i' \neq i, j$. By the choices of r_1 and δ', there are short paths from a point $\nu' \in K_i$ to ν and from ξ to a point $\xi' \in K_j$, each of duration ε_1 and cost at most 0.1ε. The path that traverses from ν' to ν, and then along μ to ξ, and thence to ξ', has cost at most $S_{[0,T]}(\mu|\nu) + 0.2\varepsilon \geq \tilde{V}(K_i, K_j)$, and so

$$S_{[0,T]}(\mu|\nu) \geq \tilde{V}(K_i, K_j) - 0.2\varepsilon. \tag{A.1}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

The same holds for any path μ of some duration T starting at ν in the r_1-neighborhood of K_i, *touching* the δ'-neighborhood of $\overline{g}_j$ at time in $[0,T]$, but not touching any of the other compact sets $K_{i'}, i' \neq i,j$.

By Lemma A.5, with the set C in place of K, a set that does not contain any ω-limit set entirely, and with $T_1 = T_0 + V_0/c$ where c, T_0 are as specified in that lemma, we obtain

$$p_\nu^{(N)}\{\tau_1 > T_1\} \leq \sup_{\nu' \in C} p_{\nu'}^{(N)}\{\tau_C > T_1\} \leq e^{-NV_0} \tag{A.2}$$

for all sufficiently large N.

Consider a trajectory μ_N with $\mu_N(0) = \nu \in \overline{[K_i]_{r_1}}$ and $\mu_N(\tau_1) \in \overline{g}_j$. There are two possibilities: (1) $\tau_1 > T_1$, or (2) $\tau_1 \leq T_1$ in which case the trajectory enters $\overline{g}_j$ in $[0,T_1]$. In this second case, with

$$\Phi_{[0,T_1],\nu}(v) := \{\mu : [0,T_1] \to \mathcal{M}_1(\mathcal{Z}) \mid S_{[0,T]}(\mu|\nu) \leq v\},$$

we have

$$\rho_{T_1}(\mu_N, \Phi_{[0,T_1],\nu}(\tilde{V}(K_i,K_j) - 0.3\varepsilon)) \geq \delta'. \tag{A.3}$$

To see this, note the conditions $\delta' < r_1$, $\tau_1 \leq T_1$, and $\mu_N(\tau_1) \in \overline{g}_j$. If μ is any trajectory satisfying $\rho_{T_1}(\mu_N, \mu) < \delta'$, then μ must hit the δ'-neighborhood of $\overline{g}_j$ without touching any of the other compact sets $K_{i'}, i' \neq i,j$. From (A.1), subtracting an extra 0.1ε, we get $S_{[0,T_1]}(\mu|\nu) > \tilde{V}(K_i,K_j) - 0.3\varepsilon$. By contraposition, under the noted conditions, any μ with $S_{[0,T_1]}(\mu|\nu) \leq \tilde{V}(K_i,K_j) - 0.3\varepsilon$ must satisfy $\rho_{T_1}(\mu_N,\mu) \geq \delta'$, and hence (A.3) follows.

Putting the two cases together, we get

$$\begin{aligned} p_\nu^{(N)}\{\mu_N(\tau_1) \in \overline{g}_j\} &\leq p_\nu^{(N)}\{\tau_1 > T_1\} \\ &\quad + p_\nu^{(N)}\{\rho_{T_1}(\mu_N, \Phi_{[0,T_1],\nu}(\tilde{V}(K_i,K_j) - 0.3\varepsilon)) \geq \delta'\} \\ &\overset{(a)}{\leq} e^{-NV_0} + \exp\{-N(\tilde{V}(K_i,K_j) - 0.3\varepsilon) + N(0.1\varepsilon)\} \\ &\overset{(b)}{\leq} \exp\{-N(\tilde{V}(K_i,K_j) - \varepsilon)\}. \end{aligned}$$

In the above sequence of inequalities, (a) holds for all sufficiently large N due to (A.2), (3.13) of Corollary 3.1, and the definition of $\Phi_{[0,T_1],\nu}$. Inequality (b) also holds for all sufficiently large N because $\tilde{V}(K_i,K_j) \leq V_0$. This proves the upper bound and completes the proof. □

Recall the definition $\mathbb{G}\{i\}$ in the paragraph preceding (2.12), the definition of $W(K_i)$ in (2.12), and the definition of $s_i, i = 1,\ldots,l$. We are now ready to state the main theorem of this appendix.

THEOREM A.1. (Freidlin and Wentzell [21, Ch.6, Theorem 4.1]). *Assume* (**A1**)-(**A3**) *and* (**B**) *hold. For any* $\varepsilon > 0$*, there is an arbitrarily small* $r_1 > 0$ *such that*

$$\exp\{-N(s_i + \varepsilon)\} \leq \wp^{(N)}\{[K_i]_{r_1}\} \leq \exp\{-N(s_i - \varepsilon)\}$$

where s_i *are defined in (2.13).*

PROOF. All the steps of the proof of [21, Ch.6, Th.4.1] hold, since the analogs of all the lemmas used in that proof have now been verified to hold. □

Acknowledgements. We thank an anonymous reviewer for suggestions that helped improve the presentation.

REFERENCES

[1] ANANTHARAM, V. (1991) "A mean field limit for a lattice caricature of dynamic routing in circuit switched networks", *Annals of Appl. Prob.* **1**, 481–503. MR1129770

[2] ANANTHARAM, V. and BENCHEKROUN, M. (1993) "A technique for computing sojourn times in large networks of interacting queues", *Probability in the Engineering and Informational Sciences* **7**, 441–464.

[3] BENAIM, M. and LE BOUDEC J.-Y. (2008) "A class of mean field interaction models for computer and communication systems", *Perform. Eval.* **65**, 823–838.

[4] BENAIM, M. and WEIBULL, J. (2003) "Deterministic approximation of stochastic evolution", *Econometrica* **71**, 873-904. MR1983230

[5] BIANCHI, G. (1998) "IEEE 802.11 - saturated throughput analysis", *IEEE Comm. Lett.* **12**, 318–320.

[6] BISWAS, A. and BORKAR, V. S. (2009) "Small noise asymptotics for invariant densities for a class of diffusions: a control theoretic view", *Journal of Mathematical Analysis and Applications* **360**, 476–484. Correction note: arXiv:1107.2277v1 MR2561245

[7] BORKAR, V. S. (1995) *Probability Theory: An Advanced Course*, Springer Verlag, New York. MR1367959

[8] BORDENAVE, C., MCDONALD, D. and PROUTIÈRE, A. (2010) "A particle system in interaction with a rapidly varying environment: mean field limits and applications", *Networks and Heterogeneous Media*, **5**(1).

[9] BORDENAVE, C., MCDONALD, D. and PROUTIÈRE, A. (2005) "Random multi-access algorithms, a mean field analysis", *Proc. 43th Allerton conference.*

[10] DAI PRA, P. and DEN HOLLANDER, F. (1996), "McKean-Vlasov limit for interacting random processes in random media", *Journal of Statistical Physics*, **84**(3/4), 735–772. MR1400186

[11] DAWSON, D. A. and GÄRTNER, J. (1987) "Large deviations from the McKean-Vlasov limit for weakly interacting diffusions", *Stochastics*, **20**, 247–308. MR0885876

[12] DAY, M. V. (1987) "Recent progress on the small parameter exit problem", *Stochastics*, **20**, 121–150. MR0877726

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

[13] Dembo, A. and Zeitouni, O. (2010) *Large Deviations Techniques and Applications*, (2nd ed., corrected printing), Springer-Verlag, Berlin Heidelberg. MR2571413

[14] Del Moral, P. and Zajic, T. (2003) "Note on the Laplace-Varadhan Integral Lemma", *Bernoulli*, **9**(1), Feb. 2003, 49–65. MR1963672

[15] Djehiche, B. and Kaj, I. (1995) "The rate function for some measure-valued jump processes", *The Annals of Probability*, **23**(3), 1414–1438. MR1349178

[16] Duffy, K. R. (2010) "Mean field Markov models of wireless local area networks", *Markov Processes and Related Fields* **16**(2), 295–328. MR2666856

[17] Ethier, S. N. and Kurtz, T. G. (2005) *Markov Processes: Characterization and Convergence* (2nd ed.), John Wiley, New York. MR0838085

[18] Feng, J. and Kurtz, T. G. (2006) *Large Deviations for Stochastic Processes*, Mathematical Surveys and Monographs, **131**. American Math. Soc., Providence, RI. MR2260560

[19] Feng, S. (1994) "Large deviations for Markov processes with mean field interaction and unbounded jumps", *Probability Th. and Related Fields* **100**, 227–252. MR1296430

[20] Feng, S. (1994) "Large deviations for empirical process of mean-field interacting particle system with unbounded jumps", *The Annals of Probability*, **22**(4), 2122–2151. MR1331217

[21] Freidlin, M. I. and Wentzell, A. D. (1998) *Random Perturbations of Dynamical Systems* (2nd ed.), Springer Verlag, Berlin–Heidelberg. MR1652127

[22] Graham, C. (2000) "Chaoticity on path space for a queueing network with selection of the shortest queue among several", *J. Appl. Probab.* **37**(1), 198–211. MR1761670

[23] Kac, M. (1956) "Foundations of kinetic theory", in '*Proc. of 3rd Berkeley Symp. on Math. Stat. and Prob.*' (J. Neyman, ed.), Uni. of California Press, 171–197. MR0084985

[24] Léonard, C. (1995) "Large deviations for long range interacting particle systems with jumps", *Annales de l'I. H. P., section B*, **31**(2), 289–323. MR1324810

[25] McKean, H. P. (1966) "A class of Markov processes associated with nonlinear parabolic equations", *Proc. Nat. Acad. Sci.* **56**, 1907–1911. MR0221595

[26] Ramaiyan, V., Kumar, A. and Altman, E. (2008) "Fixed point analysis of single cell IEEE 802.11e WLANs: uniqueness and multistability", *IEEE/ACM Trans. Network.* **16**, 1080–1093.

[27] Sharma, G., Ganesh, A. J. and Key, P. (2009) "Performance analysis of contention based medium access protocols", *IEEE Trans. Information Th.* **55**(4), 1665–1682. Also reported in *Proc. IEEE Infocom (2006)*. MR2582753

[28] Sheu, S-J. (1985) "Stochastic control and exit probabilities of jump processes", *SIAM J. Control and Optimization* **23**(2), 306–328, March 1985. MR0777462

[29] Sheu, S-J. (1986) "Asymptotic behavior of the invariant density of a diffusion Markov process with small diffusion", *SIAM J. Math. Anal.* **17**(2), 451–460. MR0826705

[30] Shwartz, A. and Weiss, A. (1995) *Large Deviations for Performance Analysis*, Chapman & Hall. MR1335456

[31] Stolyar, A. L. (1989) "Asymptotic behavior of the stationary distribution for a closed queueing system", *Problemy Peredachi Informatsii*, **25**(4), 80–92. MR1040026

[32] Sznitman, A. S. (1991) "Topics in propogation of chaos", in '*Ecole d'été de Probabilités de Saint Flour (1989)*', Springer Lecture Notes in Mathematics No. **1464**, 166–251 MR1108185

[33] Vvedenskaya, N. D. and Sukhov Yu. M. (2007), "A multi-user multiple-access system: stability and metastability", *Problemy Peredachi Informatsii* **43**, 105–111. MR2360021

Department of Electrical Engineering
Indian Institute of Technology Bombay
Powai, Mumbai 400076, India
E-mail: borkar.vs@gmail.com

Department of Electrical
Communications Engineering
Indian Institute of Science
Bangalore 560012, India
E-mail: rajeshs@ece.iisc.ernet.in

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.