Large Deviations of Jump Process Fluxes

Robert I. A. Patterson · D. R. Michiel Renger

Received: 11 January 2019 / Accepted: 25 July 2019 / Published online: 9 September 2019
© Springer Nature B.V. 2019

Abstract
We study a general class of systems of interacting particles that randomly interact to form new or different particles. In addition to the distribution of particles we consider the fluxes, defined as the rescaled number of jumps of each type that take place in a time interval. We prove a dynamic large deviations principle for the fluxes under general assumptions that include mass-action chemical kinetics. This result immediately implies a dynamic large deviations principle for the particle distribution.

Keywords Chemical reaction networks · Markov jump processes · Large deviations

Mathematics Subject Classification 2010 60F10 · 60J25 · 80A30 · 82C22

1 Introduction

Interacting stochastic particle systems with jump dynamics are widely used in all branches of the natural sciences, often in the context of many particle limits. More detailed information about the fluctuations around the limiting system can be obtained via a large deviations principle, which can be used for thermodynamic analysis. For example, in equilibrium the large deviations yield the free energy functional; whereas for reversible particle systems out of equilibrium, pathwise large deviations are related to free-energy-driven gradient flows, see [2, 35–37], and for an application to chemical reactions [34]. These gradient flows characterise dynamic behaviour even far from equilibrium.

However, a thermodynamic interpretation of non-reversible particle systems remains one of the main open problems of non-equilibrium thermodynamics. An example of the additional complexity in non-reversible systems is the existence of
non-equilibrium steady states, where microscopic particles flow round a cycle with no net effect on the distribution of particles.

This strongly suggests that one should study fluxes, not just the particle distribution, that is to say one should count the number of times each kind of jump occurs. This flux perspective is common in the study of models for particle transport, see for example \[5, 6, 9, 10, 14, 15\]. Fluxes are also implicit in some older central limit type results for chemical reaction models \[30\].

The particle transport examples just mentioned have in common that in the physically relevant limit, where space and particle number are rescaled, noise due to microscopic fluctuations is approximately Gaussian, and the large deviations rate function is consequently a quadratic action functional. By contrast, for models where the underlying state space remains discrete in the limit, for example because it is a list of chemical species, the microscopic fluctuations of the jump processes typically remain Poisson in nature, leading to large deviations rate functionals that are entropic rather than quadratic.

In this work we prove entropic large deviations for the fluxes of jump processes on a discrete state space. We consider very general assumptions on the jumps and rates, allowing for a wide range of applications, from microbiological phenomena to epidemic models. The prototypical application of these processes is as a model for chemical reactions, and we will adopt this interpretation throughout the paper, but at no point do we assume the existence of conserved quantities such as the number of atoms.

**Reacting Particle System** In the chemical interpretation we study a general network of reactions,

\[
\sum_{y \in \mathcal{Y}} \alpha_{y}^{(r)} y \xrightarrow{\tilde{k}^{(r)}} \sum_{y \in \mathcal{Y}} \beta_{y}^{(r)} y, \quad r \in \mathcal{R},
\]

where \(\mathcal{Y}\) is finite set of species, and \(\mathcal{R}\) is a finite set of reactions, and \(\tilde{k}^{(r)}\) are the corresponding reaction rates. A typical choice of jump rates when dealing with chemical reactions is \(\tilde{k}^{(r)}(c) = \text{const} \times \prod_{y \in \mathcal{Y}} c_{y}^{\alpha_{y}^{(r)}}\); this is called mass-action kinetics, but we will consider a much more general class of rates.

For example, one could have the reactions

\[
2\text{H}_2 + \text{O}_2 \xrightarrow{\tilde{k}^{(\text{fw})}} 2\text{H}_2\text{O}, \quad \text{and} \quad 2\text{H}_2\text{O} \xrightarrow{\tilde{k}^{(\text{bw})}} 2\text{H}_2 + \text{O}_2.
\]

In this case the set of species is \(\mathcal{Y} = \{\text{H}_2, \text{O}_2, \text{H}_2\text{O}\}\), the set of reactions is \(\mathcal{R} = \{\text{fw, bw}\}\), and \(\tilde{k}^{(\text{fw})}, \tilde{k}^{(\text{bw})}\) are the reaction rates that depend on the concentration of the species in \(\mathcal{Y}\). Furthermore, the species needed for the reactions can be grouped in the vectors \(\alpha^{(\text{fw})}, \alpha^{(\text{bw})} = (2, 1, 0), (0, 0, 2)\), and similarly for the species resulting from the reactions \(\beta^{(\text{fw})}, \beta^{(\text{bw})} = (0, 0, 2), (2, 1, 0)\). These vectors are called complexes or stoichiometric coefficients, the latter being Greek for “element counting”.

The reaction networks described above are commonly modelled by the following microscopic particle system, see the survey \[4\] and the references therein. If at some given time \(t\) there are \(N(t)\) particles of types \(Y_1(t), \ldots, Y_{N(t)}(t)\) in the system with fixed volume \(V\), then the empirical measure (or concentration) is defined as \(C^{(V)}(t) := V^{-1} \sum_{i=1}^{N(t)} 1_{Y_i(t)}\). With jump rate \(k^{(r,V)}(C^{(V)}(t))\), also called propensity, a
reaction $r$ occurs, causing the concentration to jump to the new state $C^{(V)}(t) + \frac{1}{V} \gamma^{(r)}$, where $\gamma^{(r)} = \beta^{(r)} - \alpha^{(r)} \in \mathbb{R}^Y$ is the effective stoichiometric vector (sometimes called state change vector) for reaction $r$ and these are collected in a matrix $\Gamma := [\gamma^{(1)}, \ldots, \gamma^{(R)}]$, which therefore maps rescaled reaction counts to changes in concentration. Since the propensities $k^{(r)}(V)$ depend on the particles through the empirical concentration only, $C^{(V)}(t)$ is a Markov jump process in $\mathbb{R}^Y$. The volume $V$ controls the order of the (changing) number of particles in the system.

A classic result \cite{28, 29} says that the empirical measure $C^{(V)}(t)$ converges as $V \rightarrow \infty$ to the solution of the reaction rate equation $\dot{c}(t) = \sum_{r \in R} \bar{k}^{(r)}(c(t)) \gamma^{(r)}$, where $V^{-1} k^{(r)} \rightarrow \bar{k}^{(r)}$ (in a way that we specify later).

**Reaction Fluxes** More information is included in the integrated empirical reaction flux, $W^{(V,r)}(t) = \frac{1}{V} \# \{ \text{reactions $r$ that occurred in time } (0, t] \}$. The pair $(C^{(V)}(t), W^{(V)}(t))$ is then a Markov process in $\mathbb{R}^Y \times \mathbb{R}^R$ with generator

$$
(Q^{(V)} f)(c, w) = \sum_{r \in R} k^{(r)}(c) \left( f(c + \frac{1}{V} \gamma^{(r)}, w + \frac{1}{V} 1_r) - f(c, w) \right).
$$

As in the Kurtz limit, this pair converges to the solution of the system of ODEs

$$
\begin{cases}
\dot{c}(t) = \Gamma \dot{w}(t) = \sum_{r \in R} \dot{w}^{(r)}(t) \gamma^{(r)}, \\
\dot{w}(t) = \bar{k}(c(t)).
\end{cases}
$$

The first equation is a continuity equation, which also holds almost surely for the microscopic pair $(C^{(V)}, W^{(V)})$, for finite $V$.

**Large Deviations** Dynamic large deviations principles for the concentrations $C^{(V)}$ have been proven in [1, 16–18, 22, 31, 33, 40, 41] under various assumptions. Large deviations for the pair $(C^{(V)}, W^{(V)})$ of concentrations and fluxes are much less well studied. Formal large deviations calculations for the reaction fluxes are found in \cite{11}, a rigorous proof for the independent case was given in \cite{39}, and a semigroup-based rigorous proof for a restrictive class of reaction fluxes can be found in \cite{27}, in particular excluding mass-action kinetics. In \cite{10} flux large deviations are rigorously derived for a single species exclusion–creation–annihilation process, which is a comparatively simple chemical model, but with the significant addition of transport. In our main result, we prove a dynamical large deviations principle for the process $(C^{(V)}, W^{(V)})$, under initial distribution $(\mu^{(V)}, \delta_0)$, where we shall assume that $\mu^{(V)}$ satisfies a large deviations principle with some rate functional $I_0$. The precise statement reads:

**Theorem 1.1** Let $\mu^{(V)}$ satisfy a large deviations principle with rate function $I_0$, and let Assumptions 2.5 on $\mu^{(V)}$ and Assumption 2.2 on $k^{(r)}$, $\bar{k}$ hold. Then the process $(C^{(V)}(t), W^{(V)}(t))_{t=0}^T$ satisfies a large-deviation principle in $BV(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$, equipped with the hybrid topology, with good rate functional $I_0 (c(0)) + J(c, w)$, where

$$
J(c, w) := \begin{cases} 
\int_0^T S \left( \dot{w}(t) \mid \bar{k}(c(t)) \right) \, dt, & (c, w) \in W^{1,1} (0, T; \mathbb{R}^Y \times \mathbb{R}^R), \text{ and } \dot{c} = \Gamma \dot{w}, \\
\infty, & \text{otherwise},
\end{cases}
$$

\(\infty\) Springer
with relative entropy

$$S(j \mid \hat{j}) := \begin{cases} \sum_{r \in R} s(j^{(r)} \mid \hat{j}^{(r)}), & \text{if } j \ll \hat{j}, \\ +\infty, & \text{otherwise,} \end{cases}$$

and

$$s(j^{(r)} \mid \hat{j}^{(r)}) := \begin{cases} j^{(r)} \log \left( \frac{j^{(r)}}{\hat{j}^{(r)}} \right) - j^{(r)} + \hat{j}^{(r)}, & j^{(r)} > 0, \\ \hat{j}, & j^{(r)} = 0, \\ \infty, & j^{(r)} < 0, \end{cases}$$

where $j \ll \hat{j}$ means that for all $r \in R$ one has $\hat{j}^{(r)} = 0 \implies j^{(r)} = 0$.

The precise set of assumptions will be stated in Section 2.2. As an immediate consequence of Theorem 1.1, we obtain the large deviations for the concentrations:

**Corollary 1.2** Let $\mu^{(V)}$ satisfy a large deviations principle with rate function $I_0$, and let Assumptions 2.5 on $\mu^{(V)}$ and Assumption 2.2 on $k^{(V)}$, $\bar{k}$ hold. Then the process $C^{(V)}$ satisfies a large deviations principle in $BV(0, T; \mathbb{R})$, equipped with the hybrid topology, with good rate functional $I_0(c(0)) + I(c)$, where

$$I(c) := \inf_{w \in W^{1,1}(0, T; \mathbb{R})} J(c, w).$$

Our main contribution is two-fold: 1) to provide a rigorous proof of general reaction flux large deviations and 2) to prove the concentration large deviations under much more general assumptions on the reaction rates $\bar{k}$ than previously known. The flux large deviations are interesting in their own right, but also serve as a proof strategy for the concentration large deviations: the flux large deviations are considerably easier to handle due to the explicitly given rate functional (1.4). The sufficient conditions on $\bar{k}$ that we use, in particular the monotonicity and what we call super-homogeneity (see Section 2.2), are to the best our knowledge new. They include mass-action kinetics but are much more general; as such our result may be applied not only to chemical reactions, but also to biochemistry, epidemiology, telecommunication, etc.

**Topology and Initial Conditions** The mathematical setting used here differs in two technical respects from that common in the literature. Firstly, we chose to work in the ‘hybrid’ topology on paths of bounded variation rather than the commonly used Skorohod J1-topology. Although we believe that our results may be adapted so as to hold in the Skorohod topology as well, the space of paths of bounded variation is much more natural for jump processes under the scaling that we consider, see for example [7] and [24] for similar topologies in the stochastics literature. The hybrid topology that we work with is very natural for the problem in the sense that the construction of a compact set in the proof of exponential tightness becomes almost straight-forward, and lower semicontinuity of the rate functional is immediate. Furthermore, this topology allows for a generalisation to infinite dimensions [23], which
we will pursue in future research. We introduce and comment on this space, topology and \( \sigma \)-algebra in more detail in Section 2.1.

The second twist is that we consider random rather than deterministic initial conditions. The reason for this is rather technical: in the proof we will need to approximate the rate functional using sufficiently regular paths, which is difficult without changing the initial condition. Instead, we assume that the random initial condition \( C^{(V)}(0) \) satisfies a large deviation principle with continuous rate functional \( I_0 \) (whereas by definition \( W^{(V)}(0) = 0 \) almost surely), which then allows us to perturb the initial condition as well. For some results we shall consider a deterministic initial condition \( C^{(V)}(0) = \tilde{c}^{(V)}(0) \) such that \( c^{(V)}(0) \to \tilde{c}(0) \in \mathbb{R}^Y \) for some limit initial condition, which will then be extended to random initial conditions via a mixture argument [8]. We stress that there is really a trade-off to be made: if we were to replace the assumption of continuous initial large deviations by a deterministic initial condition, we would need additional assumptions on the reaction rates \( \bar{k} \) in order to construct the approximating sequences that leave the initial condition invariant, see for example [1, 18].

Strategy and Overview Section 2 describes the setting of the paper: the topology used for the dynamic large deviations, the precise assumptions on the propensities, reaction rates and initial condition. We then discuss existence and convergence of the path measures, which serves as a prerequisite for the large deviations. Section 3 is dedicated to the analysis of the rate functional. Most importantly, it is shown that the rate functional has an alternative formulation as a convex dual, and that the rate functional can be approximated by curves that are sufficiently regular to be able to perform a change-of-measure. In a sense, these approximation lemmas are the core of the large deviations proof. We shall see that the fact that the rate functional has a relatively simple formulation makes these proofs rather direct (they would be much more cumbersome when proving the large deviations of the concentrations only).

Finally, Section 4 is devoted to the proof of the large deviations principle, Theorem 1.1. It will be shown that one can always construct sufficiently steep compact cones on which the path measures place all but exponentially vanishing probability. We then show the lower bound of the measures with the random initial conditions via a double tilting argument, exploiting the approximation lemmas. After this, the upper bound is proven under deterministic initial conditions, which implies the large deviations upper bound by a mixture argument. For completeness, a proof of the change-of-measure result that we use is included in the Appendix.

2 Setting

In this section we specify the setting that will be used in the paper. More specifically, we first introduce the hybrid topology used in the large deviations, and the precise assumptions on the propensities, reaction rates and initial condition that we will need. Finally, we construct the Markov process and its corresponding limit.
2.1 The Hybrid Topology

For any path \((c, w) \in L^1(0, T; \mathbb{R}^Y \times \mathbb{R}^R)\), the essential pointwise variation is

\[
\text{epvar}(c, w) := \inf_{(\tilde{c}, \tilde{w}) = (c, w)} \sup_{t-a.e.} \sum_{k=1}^{K} |(\tilde{c}(t_{k+1}), \tilde{w}(t_{k+1})) - (\tilde{c}(t_k), \tilde{w}(t_k))|,
\]

and the space of paths of bounded variation is defined as:

\[
\text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R) := \left\{ (c, w) \in L^1(0, T; \mathbb{R}^Y \times \mathbb{R}^R) : \text{epvar}(c, w) < \infty \right\}.
\]

Some key properties of paths of bounded variation include, see [3]:

(i.) Left and right limits are well-defined, and one can (and we will) always take the càdlàg version. Wherever we write \((c(0+), w(0+))\), we implicitly mean the right limit \((c(0+), w(0+))\).

(ii.) Any path \((c(t), w(t))\) of bounded variation has a measure-valued derivative \((\dot{c}(d\tau), \dot{w}(d\tau))\) in the sense that for any pair of test functions \((\phi, \psi) \in C^0_0(0, T; \mathbb{R}^Y \times \mathbb{R}^Y)\)

\[
-\int_0^T \dot{\phi}(t) \cdot c(t) \, dt - \int_0^T \dot{\psi}(t) \cdot w(t) \, dt = \int_0^T \phi(t) \cdot \dot{c}(d\tau) + \int_0^T \psi(t) \cdot \dot{w}(d\tau).
\]

Moreover, this derivative satisfies \(\|\dot{c}, \dot{w}\|_{\text{TV}} = \text{epvar}(c, w)\).

(iii.) \(\text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)\) equipped with the norm \(\|\cdot\|_{L^1} + \text{epvar}(\cdot)\) is a Banach space, and it is isometrically isomorphic to the dual of a Banach space.

Because of the last point, the space can also be equipped with a weak-* topology, which amounts to vague convergence of both the paths \((c^{(n)}, w^{(n)})\) and its derivatives \((\dot{c}, \dot{w})\). Naturally, weak-* compactness is simply characterised by norm-boundedness. Unfortunately, the weak-* topology is not metric, and hence difficult to use for stochastic analysis. Nevertheless, norm-boundedness is known to yield compactness in a slightly stronger topology [3, Prop. 3.13], which we call the hybrid topology, defined by strong \(L^1\)-convergence of the paths and vague convergence of the derivatives:

\[
(c^{(n)}, w^{(n)}) \xrightarrow{\text{hybrid}} (c, w) \iff \|(c^{(n)}, w^{(n)}) - (c, w)\|_{L^1} \to 0 \quad \text{and}
\]

\[
\int_0^T \phi(t) \cdot \dot{c}^{(n)}(d\tau) + \int_0^T \psi(t) \cdot \dot{w}^{(n)}(d\tau) \to \int_0^T \phi(t) \cdot \dot{c}(d\tau) + \int_0^T \psi(t) \cdot \dot{w}(d\tau)
\]

for all \((\phi, \psi) \in C^0_0(0, T; \mathbb{R}^Y \times \mathbb{R}^R)\).

It turns out that the hybrid topology, although not metric, is ‘perfectly normal’, which implies that the corresponding Borel \(\sigma\)-algebra behaves nicely, and all probabilistic tools that we will need are valid, see [23, Sec. 4].

---

1The hybrid topology is usually called the weak-* topology. We name it differently to distinguish it from the functional analytically defined weak-* topology. The two topologies coincide on compact sets; in infinite dimensions the distinction becomes more subtle, see [23].
2.2 The Assumptions

We now state the set of assumptions under which we will prove our main result. A central role is played by the sets of concentrations that are reachable via chemical reactions:

**Definition 2.1** (Stoichiometric simplex) For non-negative $c \in \mathbb{R}^Y$ and $\epsilon$ we define:

$$
\mathscr{S}(c) := \{ \tilde{c} = c + \Gamma w : w \in \mathbb{R}^R, w, \tilde{c} \geq 0 \} \quad \text{and} \quad \mathscr{S}_\epsilon(c) = \bigcup_{\tilde{c}} \mathscr{S}(\tilde{c} : |c - \tilde{c}| \leq \epsilon).
$$

For vectors in $\mathbb{R}^Y$ or $\mathbb{R}^R$ we write $\geq$ for the partial ordering obtained by coordinate-wise inequalities. The set of assumptions on the propensities and reaction rates are the following:

**Assumption 2.2** (Conditions on reaction rates)

(i) $k^{(r,v)}(c) = 0$ whenever $c_y < -V^{-1}y^{(r)}(c)$ for at least one $y \in Y$,

(ii) $\sup_{c \in \mathscr{S}_\epsilon(c(0))} \sum_{r \in R} \left| \frac{1}{V} k^{(v,r)}(c) - \tilde{k}^{(v)}(c) \right| \rightarrow 0$ for all $\epsilon > 0$ and $0 \leq c(0) \in \mathbb{R}^Y$,

(iii) $\bar{k} \in C^1([0, \infty)^Y; [0, \infty)^R)$,

(iv) $\sup_{c \in \mathscr{S}_\epsilon(c)} |\bar{k}(\tilde{c})| + |\nabla \bar{k}(\tilde{c})| < \infty$ for all $0 \leq c \in \mathbb{R}^Y$ and $\epsilon > 0$,

(v) $\bar{k}(\tilde{c}) \geq \bar{k}(c)$ for all $\tilde{c} \geq c \geq 0$ in $\mathbb{R}^Y$,

(vi) there exists a strictly increasing bijection $\psi : [0, 1] \rightarrow [0, 1]$ such that $\bar{k}^{(v)}(\delta c) \geq \psi(\delta) \bar{k}^{(v)}(c)$ for all $0 \leq c \in \mathbb{R}^Y$, $0 \leq \delta \leq 1$ and $r \in R$.

The first assumption is needed to make sure that the stochastic model does not allow negative concentrations. No assumptions related to boundedness or compactness of the stoichiometric simplices $\mathscr{S}(c(0))$ are required; the only assumption that is needed is (iv): that the reaction rates remain bounded on these simplices. Furthermore, the superhomogeneity assumption (vi) holds for most practical purposes, in particular for models with mass-action kinetics. We expect that the $C^1$-regularity can be relaxed to a locally Lipschitz condition, and that the monotonicity is only required in regions where the rates are small. Taken together (i) and (ii) imply that $c \geq 0$ is necessary in order to have $\bar{k}^{(v)}(c) > 0$.

**Example 2.3** Consider the simple reaction network (1.2). Microscopically, chemical reactions are usually modelled by the chemical master equation [4], which corresponds to the propensities:

$$
k^{(fw, V)}(c) = \frac{\kappa^{(fw)}}{V^2} 2 \binom{c_H V}{2} \binom{c_O V}{1} \quad \text{and} \quad k^{(bw, V)}(c) = \frac{\kappa^{(bw)}}{V} 2 \binom{c_{H2O} V}{2},
$$

for some constants $\kappa^{(fw)}, \kappa^{(bw)}$, where we implicitly set the propensities to zero when there are too few molecules present (Assumption (i)). The stoichiometric simplex is then determined by the conservation of H and O atoms: $\mathscr{S}(c) :=$
\[ \tilde{c} \in \mathbb{R}^Y : \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} (c - \tilde{c}) = 0 \]. The propensities converge in the sense of Assumption (ii)) to the mass-action reaction rates [4]:

\[
\tilde{k}^{(fw)}(c) = \kappa^{(fw)} c_{H_2}^2 c_{O_2} \quad \text{and} \quad \tilde{k}^{(bw)}(c) = \kappa^{(bw)} c_{H_2O}^2.
\tag{2.1}
\]

Then the continuity property (iii) clearly holds. Moreover, polynomial reaction rates are locally Lipschitz, hence restricted to any stoichiometric simplex \( \mathcal{S} (c(0)) \) condition (iv) is fulfilled. Naturally, the monotonicity property (v) also holds. Finally, the superhomogeneity condition (vi) is satisfied if we choose for \( \psi(\delta) = \delta^3 \); the highest degree of the polynomial reaction rates (2.1). In fact, the same arguments hold for any chemical reaction network with the chemical master equation on a microscopic level and mass-action kinetics on a macroscopical level.

**Remark 2.4** From an abstract perspective one may also consider \((c, w)\) to be the ‘concentrations’, namely of ‘species’ in \( Y \cup \mathcal{R} \). This point of view is however not especially fruitful since the \( w \) do not really behave like species; they are not bounded and the rates \( \tilde{k}^{(r)}(c, w) = \tilde{k}^{(r)}(c) \) depend on \( w \) only through \( c \) leading to a periodic structure. There seems to be little prospect of applying existing results directly in this setting.

The generality of the class of reaction rates allowed in this work comes at the price of some regularity assumptions on the initial condition:

**Assumption 2.5** (Sufficiently regular initial LDP) The initial measure \( \mu^{(v)} \) satisfies a large deviations principle in \( \mathbb{R}^Y \) with rate function \( \mathcal{I}_0 \) such that

(i) \( \mathcal{I}_0 \) is strictly convex,

(ii) \( \mathcal{I}_0 \) is continuous,

(iii) \( \mu^{(v)} \) is exponentially tight (and hence \( \mathcal{I}_0 \) is good),

(iv) \( \mathcal{I}_0 \) satisfies the conditions of Varadhan’s Integral Lemma[19, Th. 4.3.1] for linear functions, i.e. for all \( z \in \mathbb{R}^Y \),

\[
(a) \quad \lim_{M \to \infty} \limsup_{V \to \infty} \frac{1}{V} \log \int_{\{z; c(0) \geq M\}} e^{V z \cdot c(0)} \mu^{(v)} (dc(0)) = -\infty, \quad \text{or}
\]

\[
(b) \quad \limsup_{V \to \infty} \frac{1}{V} \log \int e^{V a z \cdot c(0)} \mu^{(v)} (dc(0)) < \infty \quad \text{for some } a > 1.
\]

(v) the subdifferential \( \partial \mathcal{I}_0 (c(0)) \neq \emptyset \) for all \( 0 \leq c(0) \in \mathbb{R}^Y \).

Although this list of assumptions is a bit technical, we point out that most assumptions mean that \( C^{(v)}(0) \) satisfy a ‘sufficiently nice’ large deviations principle. For thermodynamic properties, one is mostly interested in the large deviations where the process starts from the invariant measure [39, Sec. 4], which often satisfies a large deviations principle with all the necessary properties. The continuity of \( \mathcal{I}_0 \) will be exploited (and is essential) in the approximation lemmas 3.8, 3.9, 3.10 and 3.11, and the last assumption is a technical requirement that is needed to prove the large deviations lower bound for the mixture. Strictly speaking one should take the domain of \( \mathcal{I}_0 \) into account – recall that the subdifferential of a convex functional is non-empty on the interior of its domain; however to reduce clutter we simply assume (v).
convexity combined with tightness guarantees that a strong law of large numbers will hold.

### 2.3 Construction and Convergence of the Process

We denote by $\mathbb{P}^{(V)}$ the path measure of the process $(C^{(V)}(t), W^{(V)}(t))$ with jump dynamics as captured in the generator (1.3) and initial distribution $\mu^{(V)} \times \delta_0$. This is well-defined, as Assumptions 2.2(ii) and (iv) imply that the jump rates are uniformly bounded on each stoichiometric simplex $\mathcal{S}(c)$, and hence (1.3) indeed generates a Markov process on $\text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ (see [23, Sect. 4] for a discussion of the Borel $\sigma$-algebra of the hybrid topology, and related properties).

For technical reasons we shall also consider the dynamics obtained by perturbing the jump rates using exponentials of $\zeta \in C_c(0, T; \mathbb{R}^R)$, leading to the time dependent generator

$$
(Q^{(V)}_{\zeta,t}) \Phi(c, w) := \sum_{r \in R} k^{(r)}(c)e^{\zeta(r)(t)} \left[ \Phi(c + \frac{1}{V} \gamma(r), w + \frac{1}{V} 1_r) - \Phi(c, w) \right].
$$

Since the jump rates remain uniformly bounded under the perturbation, this generator also defines a path measure $\mathbb{P}^{(V)}_{\zeta}$ with initial condition $\mu^{(V)} \times \delta_0$.

In the interests of brevity we merely state the laws of large numbers for these measures, using the fact that the equations

$$
\begin{cases}
\dot{c}(t) = \Gamma \dot{w}(t), \\
\dot{w}(t) = \bar{k}(r)(c(t)) e^{\zeta(r)(t)},
\end{cases}
$$

are well posed in $W^{1,1}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ for non-negative initial data; this may be checked by a Picard–Lindelöf argument. The basic ideas of the convergence proof go back to Kurtz [28, 29].

**Proposition 2.6** Let $\zeta \in C_c(0, T; \mathbb{R}^R)$, Assumption 2.2 hold and suppose $\tilde{\mu}^{(V)}$ converges weakly\(^2\) to $\delta(\tilde{c}(0), 0)$ for some $\tilde{c}(0) \in \mathbb{R}^Y$ with $\tilde{c}(0) \geq 0$. Then the laws $\tilde{\mathbb{P}}^{(V)}_{\zeta}$ of the Markov processes with initial conditions $\tilde{\mu}^{(V)}$ and dynamics given by (2.2) converge weakly to the delta measure concentrated on the $(c, w) \in W^{1,1}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ that is the unique solution to (2.3) with initial data $(\tilde{c}(0), 0)$.

Here weak convergence is to be understood as weak convergence of probability measures on paths equipped with the hybrid topology; this follows from the (exponential) tightness that we shall prove in Section 4.1. Note that this convergence result includes the cases of random initial conditions $\tilde{\mu}^{(V)} = \mu^{(V)}$ as in Assumption 2.5, as well as the case of deterministic initial conditions $\tilde{\mu}^{(V)} = \delta(\tilde{c}^{(V)}(0), 0)$ where $\tilde{c}^{(V)}(0) \to \tilde{c}(0)$. Note also that weak convergence of probability measures on a metric space (convergence in distribution of the associated random variables) to a

---

\(^2\)Weak convergence is used here to mean the convergence of the integrals of continuous, bounded functions.
deterministic limit implies convergence in probability; this can readily be generalised to the hybrid topology on the space of bounded variation paths.

3 Analysis of the Rate Functional

A detailed knowledge of the properties of the rate function allows for a more concise presentation of the LDP, so these properties are developed here before we embark on the stochastic aspects of the proof. It will be practical to prove a dual, variational formulation of the rate functional $J$

$$\tilde{J}(c, w) = \begin{cases} \sup_{\xi \in C^1_0(0,T;\mathbb{R}^R)} G(c, w, \xi), & \text{if } \dot{c} = \Gamma \dot{w}, \\ \infty, & \text{otherwise}, \end{cases} \quad (3.1)$$

where

$$G(c, w, \xi) := \int_0^T [\xi(t) \cdot \dot{w}(dt) - H(c(t), \xi(t))] dt, \quad (3.2)$$

$$H(c, \xi) := \sum_{r \in \mathcal{R}} \bar{k}_r(c) \left( e^{c(r)} - 1 \right). \quad (3.3)$$

In Section 3.1 below we shall prove that indeed $J = \tilde{J}$. Next, in Section 3.2 we show the rate functional $J$ can be approximated by sufficiently regular paths.

Remark 3.1 $\tilde{J} : \text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R) \to [0, \infty]$ is lower semicontinuous with respect to the hybrid topology on $\text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ since for any $\xi \in C_0(0, T; \mathbb{R}^R)$ the function $(c, w) \mapsto G(c, w, \xi)$ is hybrid continuous.

Remark 3.2 One can also rewrite the rate functional as a convex dual without restricting to pairs that satisfy the continuity equation:

$$\tilde{J}(c, w) = \sup_{\xi \in C^1_0(0,T;\mathbb{R}^Y), \xi \in C^1_0(0,T;\mathbb{R}^R)} \int_0^T \xi(t) \cdot \dot{w}(dt) + \int_0^T \xi(t) \cdot \dot{c}(dt) - \int_0^T H(c(t), \xi(t)) dt. \quad (3.4)$$

A straight-forward calculation then shows that the rate functional reduces to (1.4) if the continuity equation is satisfied, and $\infty$ otherwise. The variation over the dual variable to $\dot{c}$ corresponds in some sense to zero-probability fluctuations in the continuity equation. Therefore it is more natural to omit that supremum, which also shortens notation considerably.

3.1 Equivalence of the Rate Function Representations

This section is devoted to the proof that both formulations of the rate functional coincide. For the relative entropy formulation $J$ of the rate functional, it is built into the definition (1.4) that $(c, w) \in W^{1,1}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ and $w, \dot{w} \geq 0$ for finite $J(c, w)$. The following Lemma says that the concentrations remain non-negative.
Lemma 3.3 Let \((c, w) \in BV(0, T; \mathbb{R}^J \times \mathbb{R}^R)\) and \(c(0) \geq 0\). If \(J(c, w) < \infty\) then \(c \geq 0\).

Proof Assume on the contrary that one may find \(t_1, y_1\) such that \(c(y_1(t_1)) < 0\). By definition \(J(c, w) < \infty\) implies \(c(y_1)\) (has a representative that) is absolutely continuous so one may take \(0 \leq t_2 < t_1\) such that \(0 \geq c(y_1(t_2)) > c(y_1(t_1))\). This implies the existence of \(r_1 \in \mathcal{R}\) such that \(c''(t) < 0\) and \(\int_{t_2}^{t_1} \dot{w}^{(r_1)}(s) \, ds > 0\) so (1.4) requires \(\hat{k}^{(r_1)}(c(s)) > 0\) almost everywhere in \([t_2, t_1]\). However from Assumption 2.2 parts (i) and (ii) one sees that \(\hat{k}^{(r_1)}(c(s)) = 0\) for all \(s \in [t_2, t_1]\).

In order to compare \(J\) to the variational formulation \(\tilde{J}\) we need to prove the same regularity result for \(\tilde{J}\):

Lemma 3.4 Let \((c, w) \in BV(0, T; \mathbb{R}^J \times \mathbb{R}^R)\). If \(\tilde{J}(c, w) < \infty\) then \((c, w) \in W^{1,1}(0, T; \mathbb{R}^J \times \mathbb{R}^R)\) and \(w, \dot{w}, c \geq 0\).

Proof Let \((c, w) \in BV(0, T; \mathbb{R}^J \times \mathbb{R}^R)\) and \(\tilde{J}(c, w) < \infty\); the proof is carried out in three stages:

1. \(\dot{w}\) is a non-negative measure,
2. \(\dot{w}(dr) = \dot{w}(t) \, dr\) for some non-negative density \(\dot{w} \in L^1(0, T; \mathbb{R}^R)\),
3. \(\dot{c}(t) = \Gamma \dot{w}(t)\),
4. \(c \geq 0\).

For the first point note that the existence of \(\dot{w}\) as a (signed) vector measure of finite total variation follows from [23, Th. 2.13]. Suppose now that there is some \(r \in \mathcal{R}\) and a measurable set \(A \subset (0, T)\) such that \(\dot{w}^{(r)}(A) < 0\). Using the Hahn decomposition and the regularity of Borel measures on the metric space \((0, T)\) ([12, Th. 7.1.7] or [25, Lem. 1.34]) one has the existence of a closed \(B \subset A\) with \(\dot{w}^{(r)}(B) < 0\). Define a non-positive test function \(\xi_n \in C^1_c(0, T; \mathbb{R}^R)\) by

\[
\xi_n^{(r')}(t) = \begin{cases} 
0 & r' \neq r \\
-\log |G(t)| \varphi(t) & r' = r 
\end{cases}
\]

for some \(\varphi \in C^1_c(0, T; [0, 1])\) such that \(\mathbb{1}_B \leq \varphi \leq \mathbb{1}_A\). On can now check that \(\lim_n G(c, w, \xi_n) = +\infty\), which contradicts \(\tilde{J} < \infty\) so there cannot be any \(r\) for which \(\dot{w}^{(r)}\) takes negative values.

For the absolute continuity suppose that there is an \(r \in \mathcal{R}\) and a measurable set \(A \subset (0, T)\) such that \(\dot{w}^{(r)}(A) = \delta > 0\), but \(|A| = 0\), where we write \(|\cdot|\) for Lebesgue measure. By the regularity result already mentioned in this proof we have the existence of closed sets \(F_n\) and open sets \(G_n\) such that \(F_n \subset A \subset G_n\) with \(\dot{w}^{(r)}(G_n \setminus F_n) < \frac{1}{n}\) and \(|G_n| \leq \frac{1}{n}\). Now define a non-negative test function \(\xi_n \in C^1_c(0, T; \mathbb{R}^R)\) by

\[
\xi_n^{(r')} = \begin{cases} 
0 & r' \neq r \\
-\log |G_n| \varphi(t) & r' = r 
\end{cases}
\]
for some $\varphi \in C^1_c(0, T; [0, 1])$ such that $1_{F_n} \leq \varphi \leq 1_{G_n}$ to get a contradiction as in the proof that $\dot{w} \geq 0$. The Radon-Nikodym theorem thus allows us with a little abuse of notation to write $\dot{w}^{(r)}(dt) = \dot{w}^{(r)}(t)dt$ for $\dot{w}^{(r)} \in L^1(0, T; \mathbb{R})$.

The proof that $c(t) \geq 0$ is the same as in Lemma 3.3, where now we have on the non-null set $B \subset (0, T)$,

$$\mathcal{J}(c, w) \geq \sup_{\xi^{(r)} \in C^1_b(B)} \int_B \xi^{(r)}(t) \cdot \dot{w}^{(r)}(t) - 0 = \infty.$$  

\[\square\]

**Proposition 3.5** \(\mathcal{J} = \mathcal{J}'.\)

**Proof** Let \((c, w) \in BV(0, T; \mathbb{R}^\mathcal{Y} \times \mathbb{R}^\mathcal{R})\) (possibly with \(\mathcal{J}(c, w) = \infty\)). If \((c, w) \notin W^{1, 1}(0, T; \mathbb{R}^\mathcal{Y} \times \mathbb{R}^\mathcal{R})\) or $c, w, \dot{w} \geq 0$ is violated, then by Lemmas 3.3 and 3.4 both $\mathcal{J}(c, w) = \infty = \mathcal{J}(c, w)$. Now assume that $(c, w) \in W^{1, 1}(0, T; \mathbb{R}^\mathcal{Y} \times \mathbb{R}^\mathcal{R})$ and $c, w, \dot{w} \geq 0$. We can then write $G(c, w, \zeta) = \sum_{r \in \mathcal{R}} \int_0^T g^{(r)}(c(t), \dot{w}^{(r)}(t), \xi^{(r)}(t)) dt$ where

$$g^{(r)}(c, j^{(r)}, \xi^{(r)}) := \xi^{(r)} j^{(r)} - \tilde{k}^{(r)}(c) \left( e^{\xi^{(r)}} - 1 \right).$$

We now show that

$$\mathcal{J}(c, w) = \sup_{\zeta: (0, T) \to \mathbb{R}^\mathcal{R}} G(c, w, \zeta) = \sup_{\zeta \in L^\infty(0, T; \mathbb{R}^\mathcal{R})} G(c, w, \zeta) = \sup_{\zeta \in C^1_b(0, T; \mathbb{R}^\mathcal{R})} G(c, w, \zeta) \geq \mathcal{J}(c, w). \quad (3.4)$$

The first equality in (3.4) can be calculated directly through the pointwise supremum. For the second equality, we construct, for each $t \in (0, T)$ and $r \in \mathcal{R}$, an explicit (pointwise) maximising sequence $\zeta_n^{(r)}(t)$ for $\sup_{\zeta^{(r)}} g^{(r)}(c(t), \dot{w}^{(r)}(t), \zeta^{(r)}(t))$, see Fig. 1,

$$\zeta_n^{(r)}(t) := \begin{cases} 
\left( \log \frac{w^{(r)}(t)}{\tilde{k}^{(r)}(c(t))} \wedge n \right) \vee -n, & \tilde{k}^{(r)}(c(t)) > 0 \text{ and } \dot{w}^{(r)}(t) > 0, \\
-n, & \tilde{k}^{(r)}(c(t)) > 0 \text{ and } \dot{w}^{(r)}(t) = 0, \\
n, & \tilde{k}^{(r)}(c(t)) = 0 \text{ and } \dot{w}^{(r)}(t) > 0, \\
0, & \tilde{k}^{(r)}(c(t)) = 0 \text{ and } \dot{w}^{(r)}(t) = 0.
\end{cases}$$

Then each $\zeta_n \in L^\infty(0, T; \mathbb{R}^\mathcal{R})$ and $g^{(r)}(c(t), \dot{w}^{(r)}(t), \zeta_n^{(r)}(t))$ is non-decreasing in $n$ and non-negative. Moreover, $g^{(r)}(c(t), \dot{w}^{(r)}(t), \zeta_n^{(r)}(t))$ converges pointwise in $t \in (0, T)$ and $r \in \mathcal{R}$ as $n \to \infty$ to the pointwise supremum. Hence by monotone convergence

$$\lim_{n \to \infty} \sum_{r \in \mathcal{R}} \int_0^T g^{(r)}(c(t), \dot{w}^{(r)}(t), \zeta_n^{(r)}(t)) = \sum_{r \in \mathcal{R}} \int_0^T \sup_{\zeta^{(r)}} g^{(r)}(c(t), \dot{w}^{(r)}(t), \zeta^{(r)}(t)) dt.$$

This shows that the pointwise supremum on the left of (3.4) can be taken over $L^\infty(0, T; \mathbb{R}^\mathcal{R})$.

For the third equality in (3.4) it suffices to show that for any $\zeta \in L^\infty(0, T; \mathbb{R}^\mathcal{R})$ the integrand can be approximated by a sequence in $C^2_b(0, T; \mathbb{R}^\mathcal{R})$. For an arbitrary
\( \xi \in L^\infty(0, T; \mathbb{R}^R) \) consider the convolutions with smoothing kernels \( \theta_\delta \) for \( \delta > 0 \) that weakly converges to the Dirac measure at 0 as \( \delta \to 0 \). In the convolutions we extended the function \( \xi \) to zero outside the interval \((0, T)\). Since \( \xi \in L^1(\mathbb{R}; \mathbb{R}) \) this sequence \( \xi * \theta_\delta \) converges strongly in \( L^1(\mathbb{R}; \mathbb{R}) \) to \( \xi \) as \( \delta \to 0 \), see [21, App. C.4].

By a partial converse of the Dominated Convergence Theorem [13, Th. IV.9], after passing to a subsequence \( \xi_n(t) := (\xi * \theta_{\delta_n})(t) \) converges pointwise \( t \)-almost everywhere. Then the exponential \( -\vec{k}(c(t)) \left( e^{\xi_n(t)} - 1 \right) \) integrand part of \( G(c, w, \xi_n) \) also converges pointwise for almost every \( t \). Moreover, we can bound
\[
\|\vec{k}\|_\infty \geq -\vec{k}(c(t)) \left( e^{\xi_n(t)} - 1 \right) \geq -\|\vec{k}\|_\infty e \|\xi_n\|_{L^\infty(0, T; \mathbb{R})} \geq -\|\vec{k}\|_\infty e \|\xi\|_{L^\infty(0, T; \mathbb{R})},
\]
and hence by dominated convergence
\[
-\sum_{r \in R} \int_0^T \vec{k}(c(t)) \left( e^{\xi_n(t)} - 1 \right) \, dt \to -\sum_{r \in R} \int_0^T \vec{k}(c(t)) \left( e^{\xi(t)} - 1 \right) \, dt.
\]

Clearly the linear part \( \sum_{r \in R} \int_0^T \xi_n(t) \tilde{w}(r)(t) \, dt \) of \( G(c, w, \xi_n) \) converges to \( \sum_{r \in R} \int_0^T \xi(t) \tilde{w}(r)(t) \, dt \), and so \( G(c, w, \xi_n) \to G(c, w, \xi) \). This proves the third equality in (3.4).

For the fourth equality, take any \( \xi \in C^1_b(0, T; \mathbb{R}) \), and approximate with \( \xi \eta_\delta \in C^1_c(0, T; \mathbb{R}) \) where
\[
\eta_\delta(t) := \begin{cases} 
0, & t \in (0, \delta] \cup [T - \delta, T), \\
1, & t \in [2\delta, T - 2\delta], \\
\text{smooth between 0 and 1, } t \in [\delta, 2\delta] \cup [T - 2\delta, T - \delta].
\end{cases}
\]

Then, as \( \delta \to 0 \),
\[
G(c, w, \xi \eta_\delta) = \int_0^T \tilde{w}(t) \xi(t) \eta_\delta(t) \, dt - \sum_{r \in R} \int_0^T \vec{k}(r)(c(t)) \left( e^{\xi(t)} - 1 \right) \, dt \\
\to G(c, w, \xi \eta_\delta) \to G(c, w, \xi),
\]

where for the linear part we use that $\zeta \eta_\delta \to \zeta$ weakly-* in $L^\infty(0, T; \mathbb{R}^R)$, and for the nonlinear part we use dominated convergence.

### 3.2 Approximation by Regular Curves

A common challenge in proving a large deviations lower bound for a Markov process is to approximate any curve of finite rate by curves for which one can perform a change-of-measure. In the setting of our paper, this set of sufficiently regular curves will be defined as:

$$\mathcal{A} := \left\{ (c, w) \in \text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R) \cap \text{AC}(0, T; \mathbb{R}^Y \times \mathbb{R}^R) : \zeta := \log \frac{\dot{w}}{k(c)} \in C^1_c(0, T; \mathbb{R}^R), \; \dot{c} = \Gamma \dot{w}, \; c, w, \dot{w} \geq 0, \; w(0) = 0 \right\}.$$ (3.6)

Observe that this set requires compactly supported perturbations, whereas the change-of-measure Theorem A.3 only requires boundedness. However, the compact support will be needed to control the end point in the tilting arguments, Lemmas 4.6 and 4.7. Since $\zeta$ only perturbs the dynamics, the value $\zeta(0)$ does not influence the initial condition; we just cut off the initial value as well for ease of notation.

This section is dedicated to the proof of the following approximation result, which follows by a sequence of four approximation lemmas that we prove below:

**Corollary 3.6** Let $\mu^{(V)}$ satisfy Assumption 2.5 and $\bar{k}$ satisfy Assumptions 2.2(iii), (iv), (v) and (vi). Given $(c, w) \in \text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ such that $J(c, w) < \infty$, there exists a sequence $(c_\delta, w_\delta)_\delta \subset \mathcal{A}$ such that:

1. $c_\delta(0) \to c(0)$ and $(c_\delta, w_\delta) \xrightarrow{\text{hybrid}} (c, w)$ as $\delta \to 0$.
2. $\mathcal{I}_0(c_\delta(0)) + J(c_\delta, w_\delta) \to \mathcal{I}_0(c(0)) + J(c, w)$ as $\delta \to 0$.

**Remark 3.7** Some arguments are generalisations of the previous paper for independently hopping particles, [39, Sec. 3.3]. In that setting one can construct approximation sequences without changing the initial condition. More precisely, one briefly increases and decreases the flux $\dot{w}$ a little, so that $\|\dot{w}\|_{L^1} > 0$; a simultaneous mollification on $(c, w)$ then yields a lower bounds $c$ and $\dot{w}$. The current setting is more involved, since we need to derive a lower bound on $\bar{k} \circ c$ rather than on $c$, and any such construction need to depend strongly on the interactions between different species through the operators $\bar{k}$ and $\Gamma$. Moreover, one approximation argument used in [39, Lem. 3.12] exploits the joint convexity of the function $(c, j) \mapsto s(j \mid \bar{k}(c))$, which generally fails if $\bar{k}(c)$ is not linear.

An alternative strategy (that we do not pursue) is to construct approximations that do change the initial condition, but then construct a small-time and vanished-cost connecting path between the original and the approximated initial condition [1, 18]; whether and how this can be done is a very subtle matter that requires additional
assumptions $\bar{k}$ and $\Gamma$. Our approach is more general in the sense that we do not require such assumptions, but that comes at the cost of loosening generality in the initial condition.

**Lemma 3.8** (Approximation I) Let $\mu^{(v)}$ satisfy Assumption 2.5 and $\bar{k}$ satisfy Assumptions 2.2(iii),(iv),(v) and (vi). Given $(c, w) \in \text{BV} (0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ such that $\mathcal{J} (c, w) < \infty$, there exists a sequence $(c_\delta, w_\delta)_\delta \subset \text{BV} (0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ such that:

- (a) $c_\delta (0) \to c (0)$ and $(c_\delta, w_\delta) \xrightarrow{\text{hybrid}} (c, w)$ as $\delta \to 0$,
- (b) $\mathcal{I}_0 (c_\delta (0)) + \mathcal{J} (c_\delta, w_\delta) \to \mathcal{I}_0 (c (0)) + \mathcal{J} (c, w)$ as $\delta \to 0$,
- (c) $\inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)} (c_\delta (t)) > 0$ for any $\delta > 0$.

**Proof** Without loss of generality we may assume that for each reaction $r$ there exists a concentration $\hat{c}_r (t) \in \mathbb{R}_+^y$ for which $\bar{k}^{(r)} (\hat{c}_r (t)) > 0$. Set $\hat{c} = \sum_{r \in \mathcal{R}} \hat{c}_r$, so that by the assumed monotonicity,

$$\min_{r \in \mathcal{R}} \bar{k}^{(r)} (\hat{c}) > 0.$$ 

For $\delta > 0$ define

$$c_\delta (t) \coloneqq \delta \hat{c} + (1 - \delta) c (0) + \Gamma w_\delta (t), \quad \text{and} \quad w_\delta (t) \coloneqq (1 - \delta) w (t),$$

so that $c_\delta (t) = \delta \hat{c} + (1 - \delta) c (t) \geq 0$.

The limits (a) are trivial. The lower bound (c) follows by the monotonicity and superhomogeneity Assumptions 2.2(v) and (vi):

$$\inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)} (c_\delta (t)) \geq \min_{r \in \mathcal{R}} \bar{k}^{(r)} (\delta \hat{c}) \geq \psi (\delta) \min_{r \in \mathcal{R}} \bar{k}^{(r)} (\hat{c}) > 0. \quad (3.7)$$

For the limits (b), the convergence of $\mathcal{I}_0 (c_\delta (0))$ follows by Assumption 2.5. By lower semicontinuity we only need to check $\lim \sup_{\delta \to 0} \mathcal{J} (c_\delta, w_\delta) \leq \mathcal{J} (c, w)$. Using the facts that:

- $s (x \mid y) = s (x \mid z) + x \log (z/y) - z + y$,
- $s ((1 - \delta) x \mid z) = (1 - \delta) s (x \mid z) + \delta z + (1 - \delta) x \log (1 - \delta)$, and
- $\bar{k}^{(r)} (c_\delta (t)) \geq \psi (1 - \delta) \bar{k}^{(r)} (c (t))$,

we can rewrite and estimate:

$$s \left( w^{(r)}_\delta (t) \mid \bar{k}^{(r)} (c (t)) \right)$$

$$= s \left( w^{(r)}_\delta (t) \mid \bar{k}^{(r)} (c (t)) \right) + w^{(r)}_\delta (t) \log \left( \frac{\bar{k}^{(r)} (c (t))}{\bar{k}^{(r)} (c_\delta (t))} \right) - \bar{k}^{(r)} (c (t)) + \bar{k}^{(r)} (c_\delta (t)) \quad (3.8)$$

$$\leq (1 - \delta) s \left( w^{(r)} (t) \mid \bar{k}^{(r)} (c (t)) \right) + \delta \bar{k}^{(r)} (c (t)) + (1 - \delta) w^{(r)} (t) \log (1 - \delta)$$

$$+ \bar{k}^{(r)} (c (t)) \log \left( \frac{\bar{k}^{(r)} (c (t))}{\bar{k}^{(r)} (c_\delta (t))} \right) + \left| \bar{k}^{(r)} (c (t)) - \bar{k}^{(r)} (c_\delta (t)) \right|$$

$$\leq (1 - \delta) s \left( w^{(r)} (t) \mid \bar{k}^{(r)} (c (t)) \right) + \delta \bar{k}^{(r)} (c (t))$$

$$+ (1 - \delta) w^{(r)} (t) \log \frac{1 - \delta}{\psi (1 - \delta)} + \left| \bar{k}^{(r)} (c (t)) - \bar{k}^{(r)} (c_\delta (t)) \right|. \quad (3.9)$$
Summing over \( r \) and integrating over \( t \) shows that, first for \( \delta \) sufficiently small, and then letting \( \delta \to 0 \),
\[
\mathcal{J}(c_\delta, w_\delta) \leq (1 - \delta) \mathcal{J}(c, w) + \delta T \sup_{\tilde{c} \in \mathcal{A}_\delta(c(0))} |\tilde{k}(\tilde{c})| \\
+ (1 - \delta) \log \frac{1 - \delta}{\psi(1 - \delta)} \|\dot{w}\|_{L^1} + \sqrt{|\mathcal{R}|} \text{Lip}(\tilde{k}) \|c_\delta - c\|_{L^1} \to \mathcal{J}(c, w)
\]
due to Assumption 2.2(iv).

For smoothing purposes we make use of convolutions with the heat kernels \( \theta_\epsilon : \mathbb{R} \to \mathbb{R}; t \mapsto \exp(-t^2/2\epsilon)/\sqrt{2\pi\epsilon} \).

**Lemma 3.9** (Approximation II) Let \( \mu(\cdot) \) satisfy Assumption 2.5 and \( \tilde{k} \) satisfy Assumptions 2.2(iii) and (iv). Given \( (c, w) \in \text{BV}(0, T; \mathbb{R}^J \times \mathbb{R}^R) \) such that \( \mathcal{J}(c, w) < \infty \) and \( \inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c(t)) > 0 \), there exists a sequence \( (c_\delta, w_\delta)_\delta \subset C^\infty_B(0, T; \mathbb{R}^J \times \mathbb{R}^R) \) such that:

(a) \( c_\delta(0) \to c(0) \) and \( (c_\delta, w_\delta) \xrightarrow{\text{hybrid}} (c, w) \) as \( \delta \to 0 \),

(b) \( \mathcal{J}_0(c_\delta(0)) + \mathcal{J}(c_\delta, w_\delta) \to \mathcal{J}_0(c(0)) + \mathcal{J}(c, w) \) as \( \delta \to 0 \),

(c) \( \inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c_\delta(t)) > 0 \) for any sufficiently small \( \delta > 0 \).

**Proof** Define
\[
c_\delta(t) := c(0) + (w * \theta_\delta)(0) + \Gamma w_\delta(t) \quad \text{and} \quad w_\delta(t) := (w * \theta_\delta)(t) - (w * \theta_\delta)(0),
\]
in the convolutions we extend \( w \) constantly to \( w(0) \) and \( w(T) \) outside the interval \((0, T)\). Observe that the definition is sound in the sense that \( c_\delta(t) = (c * \theta_\delta)(t) \geq 0 \) and \( w_\delta, \dot{w}_\delta \geq 0 \). Since \( (c, w) \in W^{1,1}(0, T; \mathbb{R}^J \times \mathbb{R}^R) \) by Lemma 3.4, the desired convergence (a) of the sequence can be shown by adapting the results in [21, App. C.4] to mollifiers with non-compact support. Similarly to the proof of Proposition 3.5, we pass to a (relabelled) subsequence such that in fact \( \dot{w}_\delta(t) \to \dot{w}(t) \) pointwise in almost every \( t \in (0, T) \).

We now show the lower bound (c). Since paths of bounded variation have left and right limits and \( c_\delta \) is defined by a convolution, both sets \( c|_{(0, T)} \) and \( c_\delta|_{(0, T)} \) are contained in the compact set \( K := \text{co}(c|_{[0, T]} \subset \mathbb{R}^J \), the convex hull of the range of \( c \). Hence by the continuity of \( \tilde{k} \), there exists a \( \tau > 0 \) such that for any non-negative \( \tilde{c}, \tilde{\tilde{c}} \in K \) with \( |\tilde{c} - \tilde{\tilde{c}}| < \tau \) there holds \( |\tilde{k}(\tilde{c}) - \tilde{k}(\tilde{\tilde{c}})| < \frac{1}{2} \inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c(t)) \), the right-hand side being nonzero by assumption. By Lemma 3.4 the path \( c \) is continuous (on the compact interval \([0, T]\)), and so \( c_\delta \to c \) uniformly, see [21, App. C.4], i.e. there exists a \( \bar{\delta} > 0 \) such that for any \( \delta < \bar{\delta} \) and any \( t \in [0, T] \), one has \(|c(t) - c_\delta(t)| < \tau \). Putting the ingredients together, we find, for any \( \delta < \bar{\delta} \), that \( \tilde{k}(c_\delta(t)) - \tilde{k}(c(t)) \leq \sup_{\tilde{c}, \tilde{\tilde{c}} \in K; |\tilde{c} - \tilde{\tilde{c}}| < \tau} |\tilde{k}(\tilde{c}) - \tilde{k}(\tilde{\tilde{c}})| < \frac{1}{2} \inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c(t)), \) from which we deduce the lower bound (c):
\[
\tilde{k}^{(r)}(c_\delta(t)) \geq \tilde{k}^{(r)}(c(t)) - \frac{1}{2} \inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c(\tilde{t})) \geq \frac{1}{2} \inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c(\tilde{t})). \tag{3.10}
\]

\( \square \) Springer
The convergence \( \mathcal{I}_0 (c_\delta(0)) = \mathcal{I}_0 (c(0) + (w \ast \theta_\delta)(0)) \rightarrow \mathcal{I}_0 (c(0)) \) follows by Assumption 2.5. For the convergence of \( \mathcal{J}(c_\delta, w_\delta) \), we can bound the integrand, similarly as in (3.8),

\[
0 \leq s \left( \dot{w}_\delta^{(r)}(t) \mid \bar{k}^{(r)} (c_\delta(t)) \right) \leq s \left( \dot{w}_\delta^{(r)}(t) \mid \bar{k}^{(r)} (c(t)) \right) + a \dot{w}_\delta^{(r)}(t) + \left| \bar{k}^{(r)} (c(t)) - \bar{k}^{(r)} (c_\delta(t)) \right|, \tag{3.11}
\]

where

\[
\log \left( \frac{\bar{k}^{(r)} (c(t))}{\bar{k}^{(r)} (c_\delta(t))} \right) \leq \log \left( \frac{2 \sup_{t \in (0,T)} \bar{k}(c(t))}{\inf_{t \in (0,T)} \bar{k}(c(t))} \right) =: a \in [\log 2, \infty).
\]

By the assumed continuity of the reaction rates \( s \left( \dot{w}_\delta^{(r)}(t) | k^{(r)}(c_\delta(t)) \right) \rightarrow s \left( \dot{w}^{(r)}(t) | k^{(r)}(c(t)) \right) \) pointwise in \( t \in (0, T) \). If we can prove that, after summing over \( \mathcal{R} \) and integrating over \( (0, T) \), the right-hand side in (3.11) converges to a finite integral, then \( \mathcal{J}(c_\delta, w_\delta) \rightarrow \mathcal{J}(c, w) \) by a generalisation of the Dominated Convergence Theorem, see [32, Th. 1.8 & following remark].

Naturally the last two terms converge:

\[
\sum_{r \in \mathcal{R}} \int_0^T \left[ a \dot{w}_\delta^{(r)}(t) + \left| \bar{k}^{(r)} (c(t)) - \bar{k}^{(r)} (c_\delta(t)) \right| \right] dt \rightarrow a \| \dot{w} \|_{L^1} < \infty.
\]

The convergence of the entropic part can be proven analogously to [39, Lem. 3.12]. By lower semicontinuity,

\[
\lim_{\delta \rightarrow 0} \sum_{r \in \mathcal{R}} \int_0^T s \left( \dot{w}_\delta^{(r)}(t) | \bar{k}^{(r)}(c(t)) \right) dt \geq \sum_{r \in \mathcal{R}} \int_0^T s \left( \dot{w}^{(r)}(t) | \bar{k}^{(r)}(c(t)) \right) dt.
\]

On the other hand, by Jensen’s inequality,

\[
\sum_{r \in \mathcal{R}} \int_0^T s \left( \dot{w}_\delta^{(r)}(t) | \bar{k}^{(r)}(c(t)) \right) dt \\
\leq \sum_{r \in \mathcal{R}} \int_0^T \int_{-\infty}^\infty s \left( \dot{w}^{(r)}(u) | \bar{k}^{(r)}(c(t)) \right) \theta_\delta(t - u) du dt \\
= \sum_{r \in \mathcal{R}} \int_0^T \left( \dot{w}^{(r)}(t) \log \dot{w}^{(r)}(t) \ast \theta_\delta(t) - (\dot{w}^{(r)} \ast \theta_\delta)(t) \left( 1 + \log \bar{k}^{(r)}(c(t)) \right) \right) dt \\
\rightarrow \sum_{r \in \mathcal{R}} \int_0^T s \left( \dot{w}^{(r)}(t) | \bar{k}^{(r)}(c(t)) \right) dt,
\]

again by [21, App. C.4]. Therefore the summed and integrated right-hand side of (3.11) indeed converges to a finite integral, which concludes the proof of claim (b).

\[\square\]

**Lemma 3.10** (Approximation III) Let \( \mu^{(V)} \) satisfy Assumption 2.5 and \( \bar{k} \) satisfy Assumptions 2.2(iii),(iv), (v) and (vi). Given \( (c, w) \in C_b^\infty (0, T; \mathbb{R}^Y \times \mathbb{R}^R) \) such
that $J(c, w) < \infty$ and $\inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c(t)) > 0$, there exists a sequence $(c_\delta, w_\delta)_\delta \subset C^\infty_b(0, T; \mathbb{R}^y \times \mathbb{R}^n)$ such that:

(a) $c_\delta(0) \to c(0)$ and $(c_\delta, w_\delta) \xrightarrow{\text{hybrid}} (c, w)$ as $\delta \to 0$,
(b) $\mathcal{I}_0(c_\delta(0)) + J(c_\delta, w_\delta) \to \mathcal{I}_0(c(0)) + J(c, w)$ as $\delta \to 0$,
(c) $\inf_{t \in (0, T), r \in \mathcal{R}} \tilde{k}^{(r)}(c_\delta(t)) > 0$ for any $\delta > 0$,
(d) $\inf_{t \in (0, T), r \in \mathcal{R}} \dot{w}_\delta^{(r)}(t) > 0$ for any $\delta > 0$,
(e) $\zeta_\delta := \log \frac{\dot{w}_\delta}{k(c_\delta)} \in C^1_b(0, T; \mathbb{R}^n)$.

Proof Let $0 \leq \beta^{(r)}, \alpha^{(r)} \in \mathbb{R}^y$ be the positive and negative parts of $\gamma^{(r)}$, i.e. $\gamma^{(r)} = \beta^{(r)} - \alpha^{(r)}$. For $0 < \theta < 1$ define

$$c_\delta(t) := (1 - \delta)c(0) + \delta T \sum_{r \in \mathcal{R}} \alpha^{(r)} + \Gamma w_\delta(t) \quad \text{and} \quad w_\delta(t) := (1 - \delta)w(t) + \delta t,$$

so that $c_\delta(t) = (1 - \delta)c(0) + \delta \sum_{r \in \mathcal{R}} [(T - t)\alpha^{(r)} + \theta \beta^{(r)}] \geq 0$ and $w_\delta(t) = (1 - \delta)w(t) + \delta t \geq \delta > 0$. Hence the sequence is admissible, and property (d) holds by construction. Again, the hybrid convergence (a) is trivial, and the monotonicity and superhomogeneity, Assumptions 2.2(v), (vi) imply the same estimate as (3.7), which shows that the bound (c) is indeed retained.

The convergence $\mathcal{I}_0(c_\delta(0)) \to \mathcal{I}_0(c(0))$ follows from the continuity of $\mathcal{I}_0$ and (a). As in the previous lemmas, due to lower semicontinuity it is sufficient to show $\lim \sup_{\delta \to 0} \mathcal{J}(c_\delta, w_\delta) \leq \mathcal{J}(c, w)$ in order to establish (b). We can again derive estimate (3.9), where the terms $\dot{w}_\delta^{(r)}(t) \log \tilde{k}^{(r)}(c(t))/\tilde{k}^{(r)}(c_\delta(t))$ and $\tilde{k}^{(r)}(c(t)) - \tilde{k}^{(r)}(c_\delta(t))$ can be dealt with in exactly the same manner as in the proof of Lemma 3.8. It thus remains to show convergence of the integral $\sum_{r \in \mathcal{R}} \int_0^T s \left(\dot{w}_\delta^{(r)}(t)\tilde{k}^{(r)}(c(t))\right) dt$. By the convexity of $s$ in its first argument, we get for $0 < \theta < 1$,

$$s \left(\dot{w}_\delta^{(r)}(t)\tilde{k}^{(r)}(c(t))\right) \leq (1 - \delta)s \left(\dot{w}^{(r)}(t)\tilde{k}^{(r)}(c(t))\right) + \delta s \left(1\tilde{k}^{(r)}(c(t))\right) \\
\leq s \left(\dot{w}^{(r)}(t)\tilde{k}^{(r)}(c(t))\right) - \delta \log \tilde{k}^{(r)}(c(t)) + \delta \tilde{k}^{(r)}(c(t)).$$

Since the last two terms are bounded from below and above it follows that $\lim \sup_{\delta \to 0} \mathcal{J}(c_\delta, w_\delta) \leq \mathcal{J}(c, w)$.

Finally we can prove (e) for any $\delta > 0$. Since the curve $(c_\delta, w_\delta)$ is smooth we only need to prove boundedness of the functions

$$\dot{\zeta}_\delta^{(r)}(t) = \log \frac{\dot{w}_\delta^{(r)}(t)}{k(c_\delta(t))} \quad \text{and} \quad \ddot{\zeta}_\delta^{(r)}(t) = \frac{\dot{w}_\delta^{(r)}(t)}{\dot{w}_\delta^{(r)}(t)} - \nabla_c \tilde{k}(c_\delta(t)) \cdot \dot{c}_\delta(t).$$

This follows from the boundedness away from zero of $\dot{w}_\delta$ and $\tilde{k}(c_\delta)$, together Assumption 2.2(iv). \qed

**Lemma 3.11** (Approximation IV) Let $\tilde{k}$ satisfy Assumptions 2.2(iii),(iv). Given $(c, w) \in C^\infty_b(0, T; \mathbb{R}^y \times \mathbb{R}^n)$ such that $J(c, w) < \infty$ and $\zeta := \log \dot{w}/\tilde{k}(c) \in C^1_b(0, T; \mathbb{R}^n)$, there exists a sequence $(c_\delta, w_\delta) \in W^{1,1}(0, T; \mathbb{R}^y \times \mathbb{R}^n)$ such that:

(a) $c_\delta(0) \equiv c(0)$ and $(c_\delta, w_\delta) \xrightarrow{\text{hybrid}} (c, w)$ as $\delta \to 0$, 

\[ \square \] 

Springer
(b) \( \mathcal{I}_0(c_0(0)) + \mathcal{J}(c_0, w_0) \rightarrow \mathcal{I}_0(c(0)) + \mathcal{J}(c, w) \) as \( \delta \rightarrow 0 \).

(c) \( \zeta_\delta := \log \dot{w}_\delta / \bar{k}(c_\delta) \in C^1_c(0, T; \mathbb{R}^R) \).

**Proof** Given \((c, w)\) with \( \zeta = \log \dot{w} / \bar{k} \in C^1_b(0, T; \mathbb{R}^R) \), we approximate \( \zeta_\delta := \zeta \eta_\delta \) where \( \eta_\delta \) is the usual compactly supported function (3.5). Clearly \((c, w)\) satisfies the perturbed equation below, and we define, for each \( \delta > 0 \) the path \((c_\delta, w_\delta)\) as the solution of the second perturbed equation:

\[
\begin{align*}
\dot{c}(t) &= \Gamma \dot{w}(t), \\
\dot{w}(t) &= \bar{k}(r)(c(t)) e^{\zeta(r)(t)}, \\
\dot{c}_\delta(t) &= \Gamma \dot{w}_\delta(t), \\
\dot{w}_\delta(t) &= \bar{k}(r)(c_\delta(t)) e^{\zeta_\delta(r)(t)},
\end{align*}
\]
both under the same initial conditions \((c(0), 0)\).

Let us now introduce the matrix norm,

\[
\|\Gamma\| := \max_{r \in \mathcal{R}} |\gamma(r)|. (3.12)
\]

To prove convergence (a) we first estimate for any \( 0 \leq t \leq T \),

\[
|\dot{w}_\delta(t) - \dot{w}(t)| \\
\leq \sum_{r \in \mathcal{R}} \left| \bar{k}(r)(c_\delta(t)) e^{\zeta_\delta(r)(t)} - \bar{k}(r)(c(t)) e^{\zeta(r)(t)} \right| \\
+ \sum_{r \in \mathcal{R}} \left| \bar{k}(r)(c(t)) e^{\zeta(r)(t)} - \bar{k}(r)(c_\delta(t)) e^{\zeta_\delta(r)(t)} \right| \\
\leq \text{Lip}(\bar{k}) \|\zeta\|_{L^\infty} \int_0^t |\dot{w}_\delta(\hat{t}) - \dot{w}(\hat{t})| d\hat{t} \\
+ \left( \sup_{c \in \mathcal{C}(c(0))} \sum_{r \in \mathcal{R}} \bar{k}(r)(\hat{c}) \right) \max_{r \in \mathcal{R}} \left| e^{\zeta(r)(t)} - e^{\zeta_\delta(r)(t)} \right|.
\]

From (3.5) one sees that \( \zeta_\delta(t) = \zeta(t) \) except on two intervals each with length no more than \( 2\delta \). Gronwall’s inequality yields

\[
|\dot{w}_\delta(t) - \dot{w}(t)| \leq \left( \sup_{c \in \mathcal{C}(c(0))} \sum_{r \in \mathcal{R}} \bar{k}(r)(\hat{c}) \right) \int_0^t \left| e^{\zeta(\hat{t})} - e^{\zeta_\delta(\hat{t})} \right| d\hat{t} e^{\text{Lip}(\bar{k}) \|\zeta\|_{L^\infty} t} \\
\leq 4\delta \exp(\|\zeta\|_{L^\infty})
\]

and so \( w_\delta \rightarrow w \) in \( W^{1,\infty}(0, T; \mathbb{R}^R) \), and by boundedness of the operator \( \Gamma \) also \( c_\delta \rightarrow c \) in \( W^{1,\infty}(0, T; \mathbb{R}^Y) \).

For the convergence (a) we only need to prove convergence of the dynamic rate \( \mathcal{J} \): the initial conditions are identical. Indeed, by dominated convergence together with (3.13) and \( \zeta_\delta \leq \zeta \),

\[
\mathcal{J}(c_\delta, w_\delta) = G(c_\delta, w_\delta, \zeta_\delta) \rightarrow G(c, w, \zeta) = \mathcal{J}(c, w).
\]
4 Large Deviations

We approach the proof of the main result, Theorem 1.1 with a fairly classical tilting approach with a twist. In Section 4.1 we prove exponential tightness, in Section 4.2 we prove the large deviations lower bound under initial distribution $\mu^V$, exploiting the approximation arguments from Section 3.2. In Section 4.3 we first prove the weak upper bound (i.e. on compact sets) for the conditional path measures, and then for the path measures under initial distribution $\mu(V)$ again. The exponential tightness then guarantees that the upper bound also holds on closed sets, and that the rate functional is lower semicontinuous [19, Lem. 1.2.18].

We will write $\pi_0$ for the function mapping a path to its initial value, in particular for $(c, w) \in BV(0, T; \mathbb{R}^Y \times \mathbb{R}^R)$ recalling from Section 2.1 one has $\pi_0[c] := c(0) = c(0+)$.  

4.1 Exponential Tightness

By a standard Chernoff argument, the balls 

$$B_{TV}^m := \left\{ (c, w) \in BV(0, T; \mathbb{R}^Y \times \mathbb{R}^R) : \| (\dot{c}, \dot{w}) \|_{TV} \leq m \right\}$$

(4.1)

could be used for the exponential tightness, if the initial concentration were fixed. However, in order to control the initial condition in the large deviations upper bound we work with the cones (for some initial condition $\tilde{c}(0)$):

$$C_{m, \epsilon} := \left\{ (c, w) \in B_{TV}^m : \| (\dot{c}, \dot{w}) \|_{TV} \leq m \text{ and } |(c(t), w(t)) - (\tilde{c}(0), 0)| \leq \epsilon + tm \ t-a.e. \right\}.$$

Note that within the cone $|(c(t), w(t))| \leq |\tilde{c}(0)| + \epsilon + Tm$ so that $\| (c, w) \|_{L^1} \leq T(|\tilde{c}| + \epsilon + Tm)$.

Lemma 4.1 For any $m, \epsilon > 0$ the cone $C_{m, \epsilon}$ is hybrid-compact.

Proof The cone $C_{m, \epsilon}$ is contained in the total-variation ball $B_{TV}^m$ and is clearly $L^1$-bounded, so it is relatively compact as discussed in Section 2.1. We thus need to show that $C_{m, \epsilon}$ is hybrid closed. To that aim, take a hybrid-convergent net $(c^{(\omega)}, w^{(\omega)})_\omega \subset C_{m, \epsilon}$ with limit $(c, w)$. By the weak-* lower semicontinuity of the TV-norm it follows that $\| (\dot{c}, \dot{w}) \|_{TV} \leq m$. Moreover, the pointwise bound implies that,

$$\int_A \left( |(c^{(\omega)}(t), w^{(\omega)}(t)) - (\tilde{c}(0), 0)| - \epsilon - mt \right) dt \leq 0 \quad \forall \text{ measurable } A \subset (0, T),$$

hence, after taking the limit in $\omega$,

$$\int_A \left( |(c(t), w(t)) - (\tilde{c}(0), 0)| - \epsilon - mt \right) dt \leq 0 \quad \forall \text{ measurable } A \subset (0, T),$$

which is equivalent to the pointwise bound $|(c(t), w(t)) - (\tilde{c}(0), 0)| \leq \epsilon + mt$ for the limit. \[\square\]
We first show exponential tightness for the distributions of trajectories with fixed initial conditions.

**Lemma 4.2** (Uniform Exponential tightness with fixed initial conditions) Let \( \zeta \in C_c(0, T; \mathbb{R}^R) \) and assume \( V^{-1}k^{(V)}, \bar{k} \) are bounded on stoichiometric simplices (Assumptions 2.2(ii),(iv)). Fix any non-negative convergent sequence \( V \to \infty \) \( N_0 \ni \bar{c}^{(V)}(0) \to \bar{c}(0) \in \mathbb{R}^Y \) and let \( \bar{\mathbb{P}}^{(V)}_{\zeta} \) be the law of the Markov process with generator \( \mathbb{Q}^{(V)}_{\zeta,t} \) and initial distribution \( \delta_{\bar{c}(V)(0)} \). Then for any \( \epsilon \) and \( \eta > 0 \) there exists an \( m \) (not depending on the choice \( \bar{c}(0) \)) such that

\[
\frac{1}{V} \log \bar{\mathbb{P}}^{(V)}_{\zeta} (\mathcal{C}_{m, \epsilon}) \leq -\eta.
\]

**Proof** For a \( \delta > 0 \) to be determined later, define the set (see Fig. 2):

\[
\Sigma_{\delta, \epsilon} := (c, w) \in BV(0, T; \mathbb{R}^Y \times \mathbb{R}^R) : |(c(t), w(t)) - (\bar{c}(0), 0)|
\]

\[
\leq \sigma_{\delta, \epsilon}(t) \quad t \text{ - a.e.},
\]

\[
\sigma_{\delta, \epsilon}(t) := \epsilon + \sum_{l=1}^{\lfloor T/\delta \rfloor} \frac{1}{2} \epsilon l \mathbb{1}_{[l\delta, (l+1)\delta)}(t).
\]

Then \( \Sigma_{\epsilon/(2\delta), \epsilon} \supset \Sigma_{\delta, \epsilon} \cap \mathcal{B}_{T/\delta} \) and so it suffices to prove that for any \( \eta > 0 \) we can find \( m, \delta > 0 \) such that

\[
\limsup \frac{1}{n} \log \bar{\mathbb{P}}^{(V)}_{\zeta} (\Sigma_{\delta, \epsilon}) \leq -\eta \quad \text{and} \quad (4.2)
\]

\[
\limsup \frac{1}{n} \log \bar{\mathbb{P}}^{(V)}_{\zeta} (\mathcal{B}_{m, \epsilon}^{T/\delta}) \leq -\eta. \quad (4.3)
\]

Observe that by the convergence of the initial condition, for \( V \) sufficiently large and \( \bar{\mathbb{P}}^{(V)}_{\zeta} \)-almost surely,

\[
|(c(0), w(0)) - (\bar{c}(0), 0)| \leq \frac{1}{2}\epsilon. \quad (4.4)
\]
To prove (4.3), observe that the Markov jump process $\sum_{r \in R} W^{(V,r)}(t)$ is bounded by a Poisson process $\frac{1}{\lambda} N_{V\lambda}(t)$ with $\lambda := e^{\|\Gamma\|\|\xi\|_\infty} \sup_{\Gamma/\lambda(\xi)(0)} \sum_{r \in R} (1 + \bar{k}^{(r)}) < \infty$ due to Assumptions 2.2(ii) and (iv). A standard Chernoff bound therefore yields

$$\tilde{P}^{(V)}(\sum_{r \in R} W^{(V,r)}(t) > m) \leq \exp \left( -\lambda \left( \frac{m \lambda}{1 + \|\Gamma\|} \right) \right)$$

if we choose $m := (\lambda Te + \eta)(\|\Gamma\| + 1)$.

We now prove (4.2). Because of (4.4) we may assume that for any $(c, w) \in \Sigma^{c}_{\delta, \epsilon}$ there exists an interval $(l\delta, (l + 1)\delta)$ on which the process has jumped more than $\frac{1}{2} \epsilon$. Since the norm of each jump is bounded from below by $V$ (the $W$-coordinate always jumps at least that length) we can estimate:

$$\tilde{P}^{(V)}(\sum_{r \in R} W^{(V,r)}((l + 1)\delta) - W^{(V,r)}(l\delta) > \frac{\epsilon}{2}) \leq \sum_{l=1}^{[T/\delta]} \left( \frac{1}{V} \sum_{r \in R} V W^{(V,r)}((l + 1)\delta) - V W^{(V,r)}(l\delta) > \frac{\epsilon}{2} \right)$$

where the latter is found by first applying a Chernoff bound to $\text{Prob} (aN_{V\lambda}(\delta) > aV\epsilon/2)$ for arbitrary $a > 0$ and then minimising over $a$. With the choice

$$\delta := \epsilon \exp \left( -\frac{2\eta}{\epsilon} - 1 \right),$$

we find $s(\epsilon/2|\lambda \delta) = \eta + \frac{\epsilon}{2} e^{-2\eta/\epsilon - 1} \geq \eta$ which proves (4.2).

From [8, Prop. 6] we now immediately obtain exponential tightness under the initial distribution $\mu^{(V)}$:

**Corollary 4.3** (Exponential tightness) Let $\xi \in C_{c}(0, T; \mathbb{R}^{|R|})$, assume $V^{-1}k^{(V)}, \bar{k}$ are bounded on stoichiometric simplices (Assumptions 2.2(ii),(iv) and let $\mu^{(V)}$ be exponentially tight (Assumption 2.5(iii)). Then $\tilde{P}^{(V)}(\xi)$ (under initial distribution $\mu^{(V)}$) is exponentially tight.

**Remark 4.4** The results in this section apply in particular to the unperturbed case $\xi \equiv 0$.

### 4.2 Lower Bound

We now show the large deviations lower bound, based on a tilting of both the initial condition and the dynamics. Let us briefly mention that if the approximations from
Section 3.2 could be done without changing the initial condition, then only a dynamical tilting would be needed and the large deviations with random initial condition would follow by a mixture argument as in [39, Secs. 3.3 and 3.5].

**Proposition 4.5** Let \( \mu^{(v)} \) satisfy Assumption 2.5 and \( \tilde{k} \) satisfy Assumptions 2.2. For any hybrid-open set \( \mathcal{O} \subset \text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}) \),

\[
\liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(v)}(\mathcal{O}) \geq -\inf_{(c, w) \in \mathcal{O}} I_0(c(0)) + J(c, w).
\]

**Proof** Recall the definition of the set \( \mathcal{A} \) in (3.6). Choose an arbitrary hybrid-open set \( \mathcal{O} \subset \text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}) \). From Lemma 4.6 proven below, it follows that

\[
\liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(v)}(\mathcal{O}) \geq -\inf_{(c, w) \in \mathcal{O} \cap \mathcal{A}} I_0(c(0)) + J(c, w).
\]

By Corollary 3.6 it then follows that

\[
\inf_{(c, w) \in \mathcal{O}} I_0(c(0)) + J(c, w) = \inf_{(c, w) \in \mathcal{O} \cap \mathcal{A}} I_0(c(0)) + J(c, w). \quad \Box
\]

For the lower bound it remains to prove the following:

**Lemma 4.6** Let Assumption 2.2 on the rates and Assumption 2.5 on the initial distribution hold. Let \( \mathcal{O} \subset \text{BV}(0, T; \mathbb{R}^Y \times \mathbb{R}) \) be any hybrid-open set, and \( \mathcal{A} \) be the set (3.6). Then for any \((c, w) \in \mathcal{O} \cap \mathcal{A}\),

\[
\liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(v)}(\mathcal{O}) \geq -I_0(c(0)) - J(c, w). \tag{4.5}
\]

**Proof** Take a pair \((c, w) \in \mathcal{O} \cap \mathcal{A}\), and let \( \zeta := \log \dot{w}/\dot{\tilde{k}}(c) \in C^1_c(0, T; \mathbb{R}) \) and \( z \in \partial I_0(c(0)) \), which is non-empty by Assumption 2.5. Without loss of generality we assume that \( I_0(c(0)) + J(c, w) < \infty \).

We also define a perturbed initial distribution on \( \mathbb{R}^Y \) by setting:

\[
\mu^{(v)}_\zeta(d\tilde{c}(0)) = e^{Vz \cdot \tilde{c}(0) - V \Lambda^{(v)}(z)} \mu^{(v)}(d\tilde{c}(0)), \quad \text{with}
\]

\[
\Lambda^{(v)}(z) := \frac{1}{V} \log \int e^{Vz \cdot \tilde{c}(0)} \mu^{(v)}(d\tilde{c}(0)).
\]

By Assumption 2.5, we can apply Varadhan’s Lemma [19, Th. 4.3.1], and so, combined with the assumption \( z \in \partial I_0(c(0)) \),

\[
\lim_{V \to \infty} \Lambda^{(v)}(z) = \sup_{0 \leq \tilde{c}(0) \in \mathbb{R}^Y} z \cdot \tilde{c}(0) - I_0(\tilde{c}(0)) = z \cdot c(0) - I_0(c(0)) = : \Lambda(z). \tag{4.6}
\]

Since we assumed that \( I_0(c(0)) \) is finite, it follows that (at least for sufficiently large \( V \)), the value \( \Lambda^{(v)}(z) \) is finite. Naturally \( e^{V \Lambda^{(v)}(z)} \) is simply a normalisation factor.
so that the perturbed $\mu^{(V)}_\xi$ is a probability measure. We can now define the perturbed path measure

$$\mathbb{P}^{(V)}_{\xi,z}(dc'\,dw') := \int \mathbb{P}^{(V)}_{\xi}(dc'\,dw' | c'(0) = \tilde{c}(0)) \mu^{(V)}_\xi(d\tilde{c}(0)).$$

The next step is to apply Theorem A.3 to see that

$$\log \frac{d\mathbb{P}^{(V)}_{\xi,z}}{d\mathbb{P}^{(V)}_{\xi}}(c', w') = -VG^{(V)}(c', w', \xi) - Vz \cdot c'(0) + V\Lambda^{(V)}(z). \quad (4.7)$$

When checking the applicability of the results from the Appendix one may take $K_{(c', w')} = S(c') \times \mathbb{R}^\infty$. To establish Assumption A.3 one observes that $\sup_{t \in (0, T)} |(c'(t), w'(t))|$ is bounded by $|c'(0)|$ plus a constant times the number of jumps up to time $T$, and that under $\mathbb{P}^{(V)}$ the number of jumps is stochastically dominated by a Poisson random variable with finite expectation due to Assumption 2.2 parts (ii)&(iv).

We now apply a standard tilting argument with respect to this measure. We first introduce the sets, for some arbitrary small $\epsilon > 0$ (recall that $(c, w)$ is already fixed),

$$G^\epsilon_{\xi}(c', w) := \left\{ (c', w') \in \text{BV} \left( 0, T ; \mathbb{R}^Y \times \mathbb{R}^R \right) : \right.$$  

$$\left| G(c', w', \xi) - G(c, w, \xi) \right| < \epsilon \right\}, \quad (4.8)$$

$$B^\epsilon_{\xi} = B^\epsilon_{\xi}(c(0)) := \left\{ 0 \leq c'(0) \in \mathbb{R}^Y : |c'(0) - c(0)| < \epsilon \right\}. \quad (4.9)$$

Although $G^\epsilon_{\xi}$ is not restricted to the positive cone, the probabilities are of course concentrated on non-negative concentrations and fluxes. Using (4.7),

$$\frac{1}{V} \log \mathbb{P}^{(V)}_{\xi,z}(\mathcal{O}) \geq \inf_{(c', w') \in \mathcal{O} \cap G^\epsilon_{\xi}} \left[ -z \cdot c'(0) + \Lambda^{(V)}(z) - G^{(V)}(c', w', \xi) \right]$$

$$+ \frac{1}{V} \log \mathbb{P}^{(V)}_{\xi,z}(\mathcal{O} \cap G^\epsilon_{\xi}), \quad (4.10)$$

where

$$G^{(V)}(c', w', \xi) := \int_0^T \left[ \xi(t) \cdot \dot{w}'(dt) - \sum_{r \in \mathcal{R}} \frac{1}{V} k^{(V,r)}(c'(t)) \left( \epsilon^{(r)}(t) - 1 \right) \right] dt. \quad (4.11)$$

The first term is bounded by $-z \cdot \tilde{c}(0) \geq -z \cdot c(0) - \epsilon$ by definition of $B^\epsilon_{\xi}$; for the second term we use (4.6) so that

$$\left| \Lambda^{(V)}(z) - \Lambda(z) \right| < \epsilon. \quad (4.12)$$
for $V$ sufficiently large. For the third term we estimate,

$$
\left| G^{(V)}(c', w', \xi) - G(c, w, \xi) \right| 
\leq \sup_{(c', w') \in \pi_0^{-1}[B_{\epsilon}]} \left| G^{(V)}(c', w', \xi) - G(c', w', \xi) \right| 
+ \sup_{(c', w') \in G_{\xi}^0} \left| G(c', w', \xi) - G(c, w, \xi) \right| 
\leq T(e^{\|\xi\|_{L^\infty}} + 1) \sup_{c' \in S_{\epsilon}^0(c(0))} \sum_{r \in R} \frac{1}{V} k^{(V,r)}(c') - \bar{k}^{(r)}(c') + \epsilon \leq 2\epsilon, \quad (4.13)
$$

for sufficiently large $V$ because of Assumption 2.2(ii).

To deal with the last term in (4.10), we now note that $\mu_z^{(V)}$ converges in probability to $\delta_{c(0)}$ as $V \to \infty$ since for any open ball $U$ around $c(0)$ by Assumption 2.5, in particular parts (iv)&(iii) one can apply [19, Ex. 4.3.11] to show that

$$
\lim_{V \to \infty} \frac{1}{V} \log \mu_z^{(V)}(U^c) \leq -\inf_{c(0) \in U^c} (I_0(c(0)) - z \cdot c(0) + \Lambda(z)).
$$

This is strictly negative by (4.6) and Assumption 2.5(i) yielding the convergence in probability. As a special case we have $\lim_{V \to \infty} \mu_z^{(V)}(B_{\epsilon}^c) = 0$. We can now use Proposition 2.6 (convergence in probability implies weak convergence) together with the Portemanteau Theorem to show that

$$
\liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(V)}(\xi, z \in \pi_0^{-1}[B_{\epsilon}]) \geq \liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(V)}(\xi, z \in \pi_0^{-1}[B_{\epsilon}]) - \limsup_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(V)}(\pi_0^{-1}[B_{\epsilon}]) 
\geq \liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(V)}(\xi, z \in \pi_0^{-1}[B_{\epsilon}]) - \limsup_{V \to \infty} \mu_z^{(V)}(B_{\epsilon}^c) \geq 1.
$$

The Portemanteau Theorem is applicable since $\xi \cap G_{\xi}^0$ is hybrid-open by the continuity of $(c, w) \mapsto G(c, w, \xi)$ (recall $\xi \in C_0^1(0, T; \mathbb{R}^R)$).

Putting all these estimates and convergence results together we find from (4.10) that

$$
\liminf_{V \to \infty} \frac{1}{V} \log \mathbb{P}^{(V)}(\xi) \geq -z \cdot c(0) + \Lambda(z) - G(c, w, \xi) - 4\epsilon 
= -I_0(c(0)) - J(c, w) - 4\epsilon,
$$

as $\xi$ and $z$ were chosen to make the final equality true, assuming convexity of $I_0$. This proves the claim since $\epsilon$ was arbitrary. \hfill \square

### 4.3 Upper Bound

For the upper bound we work first with a deterministic initial condition and then an argument of Biggins’ [8] to deduce the upper bound for the ‘mixture’.
Lemma 4.7 Let \( \tilde{k} \) satisfy Assumptions 2.2(iii),(ii),(iv), and suppose that the \( \tilde{c}(v)(0) \) are a non-negative convergent sequence in \( \mathbb{R}^\mathcal{Y} \) such that \( \tilde{c}(v)(0) \rightarrow \tilde{c}(0) \in \mathbb{R}^\mathcal{Y} \). Let \( \Pi^{(v)} \) be the law of the Markov process with deterministic initial condition \( \tilde{c}(v)(0) \) and the dynamics given by (1.3). Then for any hybrid-compact set \( \mathcal{K} \subset \text{BV}(0, T; \mathbb{R}^\mathcal{Y} \times \mathbb{R}^\mathcal{R}) \),

\[
\limsup_{V \to \infty} \frac{1}{V} \log \Pi^{(v)}(\mathcal{K}) \leq - \inf_{(c, w) \in \mathcal{K}} J(c, w).
\]

Proof We use an adaptation of the usual covering technique as in the proof of the Gärtner-Ellis Theorem \([19, \text{Th. 4.5.3}]\). Let \( \tilde{c}(v)(0) \) and \( \mathcal{K} \) be as stated and let \( \epsilon > 0 \).

To control the initial condition, we use the compact cones \( \mathcal{C}_{m, \epsilon} \) from the exponential tightness, Corollary 4.3. In this proof we will take \( m > 0 \) such that \( \limsup_{V \to \infty} \frac{1}{V} \log \Pi^{(v)}(\mathcal{C}_{m, \epsilon}) < -1/\epsilon \). Note that for \( V \) sufficiently large, \( \tilde{c}(v)(0) \in \mathcal{B}_\epsilon[\tilde{c}(0)] = \pi_0 \mathcal{C}_{m, \epsilon} \).

By Proposition 3.5 we can find, for any \( (c, w) \in \text{BV}(0, T; \mathbb{R}^\mathcal{Y} \times \mathbb{R}^\mathcal{R}) \) and \( \epsilon > 0 \), a \( \tilde{c}[c, w] \in C^1_c(0, T; \mathbb{R}^\mathcal{R}) \) such that \( G(c, w, \tilde{c}[c, w]) \geq J(c, w) - \epsilon \). Then the sets \( G_\epsilon[c, w] \) from (4.8) form an open covering \( \bigcup_{(c, w) \in \mathcal{K}} G_\epsilon[c, w](c, w) \supset \mathcal{K} \cap \mathcal{C}_{m, \epsilon} \), and hence there exists a finite subset \( (c^{(n)}, w^{(n)})_{n=1, \ldots, N} \subset \mathcal{K} \) such that \( \bigcup_{n=1, \ldots, N} G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}) \supset \mathcal{K} \cap \mathcal{C}_{m, \epsilon} \).

Let \( \Pi^{(v)}_\epsilon \) be the law of the Markov process with deterministic initial condition \( \tilde{c}(v)(0) \) and the dynamics given by the perturbed generator (2.2) Uniformly in \( n = 1, \ldots, N \) we find for sufficiently large \( V \)

\[
\frac{1}{V} \log \Pi^{(v)}_\epsilon(G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}))
\leq \frac{1}{V} \log \Pi^{(v)}(G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}) \cap \pi_0^{-1} \mathcal{B}_\epsilon[\tilde{c}(0)])
\leq \frac{1}{V} \log \Pi^{(v)}(G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}) \cap \pi_0^{-1} \mathcal{B}_\epsilon[\tilde{c}(0)])
\leq \frac{1}{V} \log \Pi^{(v)}_\epsilon(G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}))
\leq \frac{1}{V} \log \Pi^{(v)}_\epsilon(G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}))
\leq 0
\]

Th. A.3

\[
\sup_{(c, w) \in G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}) \cap \pi_0^{-1} \mathcal{B}_\epsilon[\tilde{c}(0)]} -G^{(v)}(c, w, \tilde{c}[c^{(n)}, w^{(n)}])
\leq \sup_{(c, w) \in G_\epsilon[c^{(n)}, w^{(n)}](c^{(n)}, w^{(n)}) \cap \pi_0^{-1} \mathcal{B}_\epsilon[\tilde{c}(0)]} -G^{(v)}(c, w, \tilde{c}[c^{(n)}, w^{(n)}]) + \epsilon
\]

(4.13)

\[
\leq -G(c, w, \tilde{c}[c^{(n)}, w^{(n)}]) + 2 \epsilon.
\]
Because of the finiteness of the covering we can now use the Laplace Principle and also exploit the control on the initial condition:

\[
\limsup_{V \to \infty} \frac{1}{V} \log \tilde{P}(V) \left( \mathcal{K}_m, \epsilon \right) \leq \limsup_{V \to \infty} \frac{1}{V} \log \left( \tilde{P}(V) \left( \mathcal{K} \cap \mathcal{C}_m, \epsilon \right) + \tilde{P}(V) \left( \mathcal{C}_n, \epsilon \right) \right)
\]

\[
\leq \max_{n=1, \ldots, N} \limsup_{V \to \infty} \frac{1}{V} \log \tilde{P}(V) \left( \mathcal{G}_n, \epsilon \left[ c^{(n)}, w^{(n)} \right] + \mathcal{C}_m, \epsilon \right) \vee - \frac{1}{\epsilon}
\]

\[
\leq \max_{n=1, \ldots, N} \left( - G(c^{(n)}, w^{(n)}, \zeta), + 2 \epsilon \right) \vee - \frac{1}{\epsilon}
\]

\[
\leq \max_{n=1, \ldots, N} \left( - G(c^{(n)}, w^{(n)}) + 3 \epsilon \right) \vee - \frac{1}{\epsilon}
\]

\[
\leq \left( - \inf_{(c, w) \in \mathcal{X} \cap \mathcal{C}_m, \epsilon} J(c, w) + 3 \epsilon \right) \vee - \frac{1}{\epsilon}
\]

\[
\leq \left( - \inf_{(c, w) \in \mathcal{X} \cap \mathcal{C}_m, \epsilon(c(0), \epsilon)} J(c, w) + 3 \epsilon \right) \vee - \frac{1}{\epsilon}
\]

This proves the claim as \( \epsilon \) was chosen arbitrarily.

We can now deduce the large deviations upper bound for the mixture:

**Corollary 4.8** Let \( \mu^{(V)} \) satisfy Assumption 2.5 and \( \tilde{k} \) satisfy Assumptions 2.2. For any hybrid-compact set \( \mathcal{K} \subset \text{BV} \left( 0, T; \mathbb{R}^N \times \mathbb{R}^R \right) \),

\[
\limsup_{V \to \infty} \frac{1}{V} \log P(V) (\mathcal{K}) \leq - \inf_{(c, w) \in \mathcal{X}} J_0(c(0)) + J(c, w).
\]

**Proof** By Assumption 2.5 and Lemma 4.7 one can apply [8, Lemma 12]. Although that result in [8] is stated under the assumption of the full large deviations principle, the proof only uses the upper bound, which we proved in Lemma 4.7.

**Acknowledgements** This research has been funded by the Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 “Scaling Cascades in Complex Systems”, Project C08.

**Appendix: A change-of-measure result for linear test functionals on jump processes**

Changes of measure are central to the proof of the large deviations principle presented in this work. This appendix arose out of the need to clarify under exactly what technical conditions [26, Appendix 1, Prop. 7.3] could be adapted to the setting of the present work, in particular so that functions of the form \( x \mapsto \zeta \cdot x \) could be used since these are not bounded functions (although they are bounded linear operators). This boundedness restriction is avoided in [38], but functions used in the change of measure are no longer time dependent and the conditions are less explicit. Here the
aim is to include unbounded, time dependent functions in the change of measure formula, but to give relatively explicit, sufficient conditions that can easily be checked using the model assumptions from the main part of the paper. In this endeavour the results are restricted to pure jump processes.

Let \( \mathcal{X} \) be a Banach space, \( T \in (0, \infty] \) and \( (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T)}) \) be a filtered probability space with canonical random variable \( X : \Omega \rightarrow \Omega \), where

- \( \Omega \) is a subset of the càdlàg functions \([0, T) \rightarrow \mathcal{X}\), with the convention \( f(T) := f(T^-) \) if \( T < \infty \),
- \( \mathcal{F} \) is the Borel \( \sigma \)-algebra generated by a separable topology on \( \Omega \) and equal to the \( \sigma \)-algebra generated by the time evaluation functions \( X \mapsto X(t) \).

Note that \( T = \infty \) is allowed for now. The application in this paper is to the case \( \Omega = \text{BV}(0, T, \mathbb{R}^Y \times \mathbb{R}^R) \) with the hybrid topology, but this is not a necessary assumption.

We define the jump process through a given family of jump kernels \( (\alpha_t(x, \cdot)_{t \in [0,T], x \in \mathcal{X}} \) where \( \alpha_t(x, A) \) is the instantaneous jump rate at time \( t \) from \( x \in \mathcal{X} \) into a measurable set \( A \subset \mathcal{X} \), together with a given initial distribution \( \mu \). Let \( \mathbb{P} \) be the law of this process, a probability measure on \( (\Omega, \mathcal{F}) \) and \( \mathbb{E} \) the associated expectation operator.

We now define a class of test functions for which the associated propagators (a two-parameter semigroup of linear operators) are well-defined. To construct this set we will assume that there exists a family of measurable (not necessarily compact or bounded) subsets \((K_x)_{x}\) of \( \mathcal{X} \) such that for all \( x \in \mathcal{X} \):

- \( x \in K_x \) and \( \bigcup_{y \in K_x} K_y = K_x \),
- \( \int_0^T \sup_{y \in K_x} \alpha_t(y, X)dt < \infty \),
- \( \sup_{t \in [0,T)} \sup_{y \in K_x} \alpha_t(y, \mathcal{X} \setminus K_x) = 0 \).

This expresses the idea that the process started from \( x \) can neither explode nor leave \( K_x \). Then the propagators \((P_{s,t}f)(x) := \mathbb{E}[f(X(t)) | X(s) = x] \) preserve the set

\[
B_K(\mathcal{X}) := \left\{ f : \mathcal{X} \rightarrow \mathbb{R} \text{ measurable, such that } \forall x \in \mathcal{X} \sup_{y \in K_x} |f(y)| < \infty \right\},
\]

and satisfy \( \frac{d}{ds}(P_{s,t}f)(x) = -(Q_t P_{s,t}f)(x) \) with (time-dependent) generator

\[
(Q_t f)(x) := \int_\mathcal{X} [f(y) - f(x)] \alpha_t(x, dy).
\]

We now make three additional assumptions under which the change-of-measure formula holds.

- there is a \( \gamma > 0 \) such that \( |y - x| \leq \gamma \) for all \( x \in \mathcal{X}, t \in (0, T) \), and \( \alpha_t(x, \cdot) - \text{a.e.} y, \) \hspace{1cm} (A.1)
- \( \lim_{n \to \infty} \mathbb{P}(\tau_n < t) = 0 \) for all \( t \in (0, T), n \in \mathbb{N} \), where \( \tau_n := \inf \{ t : \alpha_t(X(t), \mathcal{X}) \geq n \} \), \hspace{1cm} (A.2)
\[ Z^\beta(t) := \exp (\beta|X(0)|) + \exp (\beta|X(t)|) \\
+ \int_0^t \exp (\beta|X(s)|) \beta\gamma\alpha_t (X(s), X') ds. \]

The next result is a variation on [26, Appendix 1, Lem. 5.1]:

**Proposition A.1** Let \( f : [0, T) \times \mathcal{X} \to \mathbb{R} \) be bounded, absolutely continuous in \( t \) and measurable in \( x \), with measurable, uniformly bounded derivative \( \partial_t f(t, x) \). Then under Assumptions (A.2) & (A.1),

\[ M^f(t) := f(t, X(t)) - f(0, X(0)) - \int_0^t ((\partial_s + Q_s) f) (s, X(s)) ds \]

is a Martingale in the filtration \((\mathcal{F}_t)_{t \geq 0}\) generated by \( X(t) \).

**Proof** In the case that \( f \) does not depend on time and \( \sup_{t,x} \alpha_t (x, \mathcal{X}) < \infty \) the result follows from [20, Ch. 4 Sect. 7]. The additional term \( \partial_s \) is added for time-dependent test functions due to a chain rule. By approximating by the process stopped at \( \tau_n \) and using Assumptions (A.2)&(A.3) one can remove the boundedness assumption on \( \alpha \).

**Lemma A.2** Under Assumptions (A.1), (A.2) & (A.3), the conclusion of Proposition A.1 is valid when \( f(t, x) = \zeta(t) \cdot x \) and when \( f(t, x) = e^{\zeta(t) \cdot x} \), in both cases for \( \zeta \in C^1_b([0, T); \mathcal{X}^*) \) where \( \mathcal{X}^* \) is the Banach dual of \( \mathcal{X} \).

**Proof** The exponential case is proved here; the linear case is similar. Let \( \theta_n \in C^\infty(\mathbb{R}) \) be such that \( \theta_n(y) = y \) for \( y \leq n \), \( \theta_n \leq n + 1 \) and \( 0 \leq \theta'_n \leq 1 \). Take an arbitrary \( \zeta \in C^1_b([0, \infty); \mathcal{X}^*) \) and set \( f_n(t, x) = \exp(\theta_n(\zeta(t) \cdot x)) \) so that Proposition A.1 can be applied to \( M^f_n(t) \).

It follows from the definitions that for all \( t \) and \( x \),

\[ \lim_n f_n(t, x) = f(t, x) \quad \text{and} \quad \lim_n \partial_t f_n(t, x) = \partial_t f(t, x). \]

Because of Assumption (A.1) \( Q_t f \) is well defined and one can prove by dominated convergence that \( \lim_n (Q_t f_n)(t, x) = (Q_t f)(t, x) \) for all \( t, x \). Preparatory to further applications of dominated convergence we estimate

\[ f_n(t, x) \leq \exp (\|\zeta\|_{\infty} |x|), \]
\[ |\partial_t f_n(t, x)| \leq \exp (\|\zeta\|_{\infty} |x|) \|\dot{\zeta}\|_{\infty} |x|, \quad \text{and} \]
\[ |(Q_t f_n)(t, x)| \leq \exp (\|\zeta\|_{\infty} |x|) (\exp (\|\zeta\|_{\infty} y) + 1) \alpha_t (x, \mathcal{X}). \]

With these estimates and Assumption (A.3) one checks \( \lim_n M^{f_n}(t) = M^f(t) \) almost surely. Again using Assumption (A.3) one can find a \( \beta > 0 \) such that
\[ |M^{f_n}(t)| \leq Z^\beta(t) \] almost surely. By the conditional expectation form of the dominated convergence theorem, for \( s < t \),
\[ M^f(s) = \lim_n M^{f_n}(s) = \lim_n \mathbb{E} \left[ M^{f_n}(t) | \mathcal{F}_s \right] = \mathbb{E} \left[ \lim_n M^{f_n}(t) | \mathcal{F}_s \right] = \mathbb{E} \left[ M^f(t) | \mathcal{F}_s \right]. \]

Finally, for the exponential change of measure we will need a bounded time interval.

**Theorem A.3** Let \( T < \infty \), \( \zeta \in \mathbb{C}_{b}^1(0, T; \mathbb{R}^R) \), and let Assumptions (A.1), (A.2) and (A.3) all hold. Suppose \( \mathbb{P}_\zeta \) is the law of some process with paths in \( \Omega \) and having initial distribution \( \mu \). Under \( \mathbb{P}_\zeta \), \( X \) is a Markov process with generator
\[ (Q_{\zeta,t} f)(x) = \int \alpha_t(x, dy) \]
if and only if
\[ \log \frac{d\mathbb{P}_\zeta}{d\mathbb{P}}(X) = \zeta(T) \cdot X(T) - \zeta(0) \cdot X(0) - \int_0^T e^{-\zeta(t) \cdot X(t)} \partial_t + Q_\zeta \cdot e^{\zeta(t) \cdot X(t)} dt. \]
(A.4)

**Proof** We only need to show the direction “\( \Leftarrow \)”; the converse then follows immediately from the uniqueness of the generator. To this end define \( \widehat{\mathbb{P}}_\zeta \) by (A.4) and let the associated expectation operator be \( \widehat{\mathbb{E}}_\zeta \). We sketch a number of steps, similar to [26, Appendix 1, Sect. 7] and [38], by which it is shown that under \( \widehat{\mathbb{P}}_\zeta \) \( X \) is Markov with generator \( Q_{\zeta,t} \).

1. Define for \( t \in (0, T) \), the process
   \[ E(t) := \exp \left( \zeta(t) \cdot X(t) - \zeta(0) \cdot X(0) - \int_0^t e^{-\zeta(s) \cdot X(s)} \partial_s + Q_\zeta \cdot e^{\zeta(s) \cdot X(s)} ds \right) \]
   and recall \( E(T) = \lim_{t \nearrow T} E(t) \). By Lemma A.2 above, \( E(t) \) is a strictly positive, mean-one \( \mathbb{P} \)-Martingale. One then shows that \( \frac{d\mathbb{P}_\zeta}{d\mathbb{P}} \bigg|_{\mathcal{F}_t} = E(t) \) and \( \frac{d\mathbb{P}_\zeta}{d\mathbb{P}} \bigg|_{\mathcal{F}_T} = \frac{1}{E(T)} \).

2. For any \( Y \in L^1(\Omega, \mathcal{F}) \), using the definition of conditional expectation and the results from the previous point, it follows that \( \widehat{\mathbb{E}}_\zeta \left[ Y | \mathcal{F}_t \right] = \mathbb{E} \left[ YE(T) / E(t) | \mathcal{F}_t \right] \).

3. Next one can use the result from point 2 to show via conditional expectations under \( \mathbb{P} \) and the \( \mathbb{P} \)-Markov property that for \( t \geq s \) and any bounded and measurable \( f : \mathcal{X} \rightarrow \mathbb{R} \), we have
   \[ \widehat{\mathbb{E}}_\zeta \left[ f(X(t)) | \mathcal{F}_s \right] = \widehat{\mathbb{E}}_\zeta \left[ f(X(t)) | \sigma(X(s)) \right], \]
   and so \( X \) is \( \widehat{\mathbb{P}}_\zeta \)-Markov.

4. Finally, the propagators \( (P_{\zeta,s,t} f)(x) := \widehat{\mathbb{E}}_\zeta \left[ f(X(t)) | X(s) = x \right] \) then satisfy \( \frac{d}{ds} (P_{\zeta,s,t} f)(x) = - (Q_{\zeta,s} P_{\zeta,s,t} f)(x) \). This implies that under \( \widehat{\mathbb{P}}_\zeta \), \( X \) has the same finite dimensional distributions as the process with generator \( Q_{\zeta,t} \) and thus \( \widehat{\mathbb{P}}_\zeta = \mathbb{P}_\zeta \).
### References

1. Agazzi, A., Dembo, A., Eckmann, J.-P.: Large deviations theory for Markov jump models of chemical reaction networks. Ann. Appl. Probab. **28**(3), 1821–1855 (2018)
2. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. Commun. Math. Phys. **307**(3), 791–815 (2011)
3. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Oxford University Press, Oxford (2000)
4. Anderson, D.F., Kurtz, T.G.: Design and analysis of biomolecular circuits: engineering approaches to systems and synthetic biology, chapter Continuous time Markov chain models for chemical reaction networks, pp 1–44. Springer (2011)
5. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Current fluctuations in stochastic lattice gases. Phys. Rev. Lett. **94**(3), 030601 (2005)
6. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Non equilibrium current fluctuations in stochastic lattice gases. J. Stat. Phys. **123**(2), 237–276 (2006)
7. Bertini, L., Faggionato, A., Gabrielli, D.: Large deviations of the empirical flow for continuous time Markov chains Annales de l’Institut Henri poincaré. Probabilités et Statistiques **51**(3), 867–900 (2015)
8. Biggins, J.D.: Large deviations for mixtures. Electron. Commun. Probab. **9**, 60–71 (2004)
9. Bodineau, T., Lagouge, M.: Current large deviations in a driven dissipative model. J. Stat. Phys. **139**(2), 201–218 (2010)
10. Bodineau, T., Lagouge, M.: Large deviations of the empirical currents for a boundary-driven reaction diffusion model. Ann. Appl Probab. **22**(6), 2282–2319 (2012)
11. Baiesi, M., Maes, C., Netočný, K.: Computation of current cumulants for small nonequilibrium systems. J. Stat. Phys. **135**(1), 57–75 (2009)
12. Bogachev, V.I.: Measure theory, vol. I and II. Springer, Berlin (2007)
13. Brézis, H.: Analyse fonctionelle, Théorie et applications (In French, English translation available). Mason, Paris, France (1983)
14. Derrida, B., Douçot, B., Roche, P.-E.: Current fluctuations in the one-dimensional symmetric exclusion process with open boundaries. J. Stat. Phys. **115**(3), 717–748 (2004)
15. Derrida, B.: Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. J. Stat. Mech: Theory Exp. **2007**(7), P07023 (2007)
16. Dupuis, P., Ellis, R.S., Weiss, A.: Large deviation principle for finite-state mean field interacting particle systems with unbounded jumps. Ann. Probab. **22**(4), 1679–2274 (1994)
17. Evans, L.C.: Partial differential equations. American Mathematical Society, Providence (2002)
18. Feng, S.: Large deviations for empirical process of mean-field interacting particle system with unbounded jumps. Ann. Probab. **22**(4), 1679–2274 (1994)
19. Heida, M., Patterson, R.I.A., Renger, D.R.M.: Topologies and measures on the space of functions of bounded variation taking values in a Banach or metric space. Journal of Evolution Equations First online: 14 September (2018)
20. Jakubowski, A.: A non-Skorohod topology on the Skorohod space. Electron. J. Probab. **2**(4), 1–21 (1997)
21. Kallenberg, O. Foundations of modern probability. Probability and its Applications, second edn. Springer, New York (2002)
22. Kipnis, C., Landim, C.: Scaling limits of interacting particle systems. Springer, Berlin (1999)
23. Kraaij, R.: Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton-Jacobi equation. arXiv:1711.00274 (2017)
24. Kurtz, T.G.: Solutions of ordinary differential equations as limits of pure jump processes. J. Appl. Probab. **7**(1), 49–58 (1970)
29. Kurtz, T.G.: The relationship between stochastic and deterministic models for chemical reactions. J. Chem. Phys. 57(7), 2976–2978 (1972)
30. Kurtz, T.G.: Strong approximation theorems for density dependent markov chains. Stoch. Process. Appl. 6(3), 223–240 (1978)
31. Léonard, C.: Large deviations for long range interacting particle systems with jumps. Annales de l’Institut Henri Poincaré, section B 31(2), 289–323 (1995)
32. Lieb, E.H., Loss, M. Analysis, 2nd edn. American Mathematical Society, Providence (2001)
33. Li, T., Lin, F.: Large deviations for two scale chemical kinetic processes. Commun. Math. Sci. 15(1), 123–163 (2015)
34. Mielke, A., Patterson, R., Peletier, M.A., Renger, D.R.M.: Non-equilibrium thermodynamical principles for chemical reactions with mass-action kinetics. SIAM J. Appl. Math. 77(4), 1562–1585 (2017)
35. Mielke, A., Peletier, M.A., Renger, D.R.M.: On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. Potential Anal. 41(4), 1293–1327 (2014)
36. Onsager, L.: Reciprocal relations in irreversible processes I. Phys. Rev. 37(4), 405–426 (Feb 1931)
37. Onsager, L.: Reciprocal relations in irreversible processes II. Phys. Rev. 38(12), 2265–2279 (Dec 1931)
38. Palmowski, Z., Rolski, T.: A technique for exponential change of measure for Markov processes. Bernoulli 12(6), 767–785 (2002)
39. Renger, D.R.M.: Flux large deviations of independent and reacting particle systems, with implications for macroscopic fluctuation theory. J. Stat. Phys. 172(5), 1291–1326 (2018)
40. Shwartz, A., Weiss, A.: Large deviations for performance analysis: queues, communications, and computing. Chapman & Hall, London (1995)
41. Shwartz, A., Weiss, A.: Large deviations with diminishing rates. Math. Oper. Res. 30(2), 281–310 (2005)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.