Individual Molecules Dynamics in Reaction Network Models

TRACKABLE SPECIES DYNAMICS IN REACTION NETWORK MODELS*

DANIELE CAPPELLETTI† AND GRZEGORZ REMPALA‡

Abstract. In a stochastic reaction network setting we define a subset of species as 'trackable' if we can consistently follow the fate of its individual molecules. We show that using the classical large volume limit results, we may approximate the dynamics of a single molecule of trackable species in a simple and computationally efficient way. We give examples on how this approach may be used to obtain various characteristics of single-molecule dynamics (for instance, the distribution of the number of infections in a single individual in the course of an epidemic or the activity time of a single enzyme molecule). Moreover, we show how to approximate the overall dynamics of trackable species in the full system with a collection of independent single-molecule trajectories, and give explicit bounds for the approximation error in terms of the reaction rates. This approximation, which is well defined for all times, leads to an efficient and fully parallelizable simulation technique for which we provide some numerical examples.

Key words. Single-molecule dynamics, mathematical epidemiology, law of large numbers, Poisson process representation, stochastic approximation, dynamic survival analysis, Skorohod topology

AMS subject classifications. 60J28, 92C40, 92C42, 60F05

1. Introduction.

Recent advances in modeling molecular systems, especially our improved ability to track individual proteins, and the deluge of data from the observations of both molecular and macro system (think, for instance, of the ongoing COVID-19 pandemic), have created new scientific challenges of considering models of very high resolution where the dynamics of a specific bio-molecule or a particular individual are of interest. In general, such 'agent-based' models are known to be computationally very costly, due to complex stochastic dynamics and highly noisy behavior of individual agents. However, it appears that, at least in some cases, simple yet satisfactory approximation of individual molecular trajectory may be directly inferred with the help of a classical approach of stochastic chemical kinetics that assumes that all molecules or individuals are indistinguishable and consequently focuses only on their aggregated counts. As an example of one such idea, originally proposed in [7] and latter expanded in [15], consider the stochastic 'susceptible-infected' (SI) chemical reaction network where a collection of $m+n$ molecules (or individuals) is partitioned into two types: susceptible (S) and infected (I) with initially n being of type S and remaining m of type I. The stochastic network evolves in time according to a Markov jump process that counts the 'infection events', that is, the interactions of one molecule of I-type with one molecule of S-type. Each such interaction creates a new molecule of I-type and removes one of S-type (equivalently, a molecule changes its type from S to I). Accordingly, in the reaction network notation described below in Section 2.2 this model may be represented as

$$S + I \longrightarrow 2I. \tag{1.1}$$

If the rate constant of the above reaction is β/n and we assume the usual mass action kinetics [6], it is well know that the above stochastic reaction network satisfies the

*Submitted to the editors DATE.

Funding: DC was supported by the MIUR grant Dipartimenti di Eccellenza 2018-2022 (E11G18000350001). GAR was supported by the National Sciences Foundation grant (DMS-1853587).

†Politecnico di Torino, Torino, Italy (daniele.cappelletti@polito.it).

‡The Ohio State University, Columbus, Ohio, USA (rempala.3@osu.edu).

law of large numbers (admits the fluid limit approximation), in the sense that as $m, n \to \infty$ and $m/n \to \rho > 0$ the surviving proportion s_t of the S-type molecules follows the logistic equation that may be written in the form

$$-\dot{s}_t/s_t = \beta(1 + \rho - s_t) \qquad s_t(0) = 1. \tag{1.2}$$

Consequently, for $t \geq 0$ we have

$$s_t = \frac{1+\rho}{1+\rho\exp(\beta(1+\rho)t)}. \tag{1.3}$$

Thus, from the viewpoint of a single, randomly selected S-type molecule, the quantity s_t defines a *survival function* describing the limiting *probability* of surviving beyond time $t > 0$. The formula (1.3) led to the method of approximating the distribution of surviving molecules of S dubbed 'dynamic survival analysis' (DSA) described in [15] and applied recently to epidemic modeling [8, 9, 14, 21, 23]. The idea is further illustrated in Figure 1 where the average of the Markov process (1.1) is compared to the average of independent realizations of single molecule dynamics (which may be efficiently calculated using modern parallel computing capabilities). Note (1.2) may be also interpreted as the equation for the *hazard function* associated with s_t. This fact has some relevance for statistical inference, and is further exploited, for instance, in [9, 15].

Beyond the simple SI example, the DSA approach has been applied (mostly in the context of epidemics) only to a handful of reaction networks representing the so-called one-directional transfer models [7]. In all such networks individual molecules can only change their state in an ordered way, hence previously visited states are no longer attainable (for instance in the SI model a molecule of S-type can only change into I-type, but not vice-versa).

In the current paper we formally expand the survival function approach for tracking the fate of individual molecules to a much broader class of networks, including those where molecules can return to their previous stages. A simple example is obtained by augmenting the SI network with the additional reaction $I \to S$, leading to the so-called SIS model (which is of interest in epidemiology) discussed in more detail in Example 4.3 below. To establish our results for such networks, we explore a different representation of the DSA approximation, which does not explicitly involve the survival function. Continuing with the SI model example, denote by $Y^i(t)$ the binary variable that takes value 1 or 0 according to whether i-th molecule is of type S or I. The limit dynamics of an i-th individual molecule (initially of type S) is then given by

$$Y^i(t) = 1 - N^i\left(\beta\int_0^t Y^i(u)(1+\rho-s_u)du\right)$$

where N^i is the unit Poisson process tracking the transition of the i-th molecule from S-type to I-type. Note that the argument of N^i is the cumulative hazard corresponding to integral of the right-hand side of (1.2) (see [15]). Such Poisson process representation is of course completely equivalent to simply having the time of switching of the i-th molecule from S to I follow the survival function (1.3), but it allows for a description of more complex scenarios than one-directional transfer models. For example, we will prove below that the limit dynamics of a single molecule

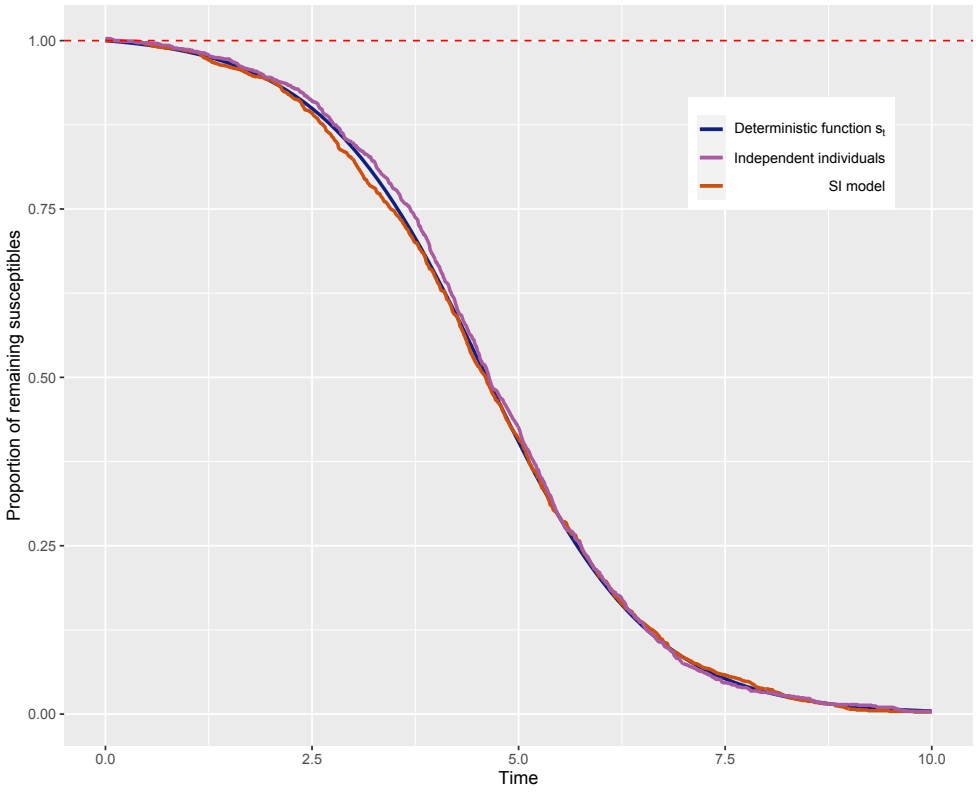

FIG. 1. ***Survival approximation in the SI model.*** *The empirical trajectory of the proportion of the remaining S molecules in the SI model described in (1.1) as compared to the deterministic function s_t defined in (1.2) and the average of $1,000$ independent single trajectories of individuals who become infected according to s_t. For the simulation we considered $n = 1,000$, $m = 10$, $\beta = 1$, and $\rho = 0.01$.*

in the SIS model can be written as

$$Y^i(t) = 1 - N_1^i\left(\beta\int_0^t Y^i(u)(1+\rho-s_u)du\right) + N_2^i\left(\kappa\int_0^t (1-Y^i(u))du\right)$$

for independent and identically distributed unit-rate Poisson processes N_1^i and N_2^i. Here, κ is the rate constant of the reaction $I \to S$.

In this work we study the Poisson process representation of the DSA approximation and give conditions under which it describes a single-molecule trajectory of the original network. In particular, we explicitly derive error bounds of the DSA approximation, in terms of the underlying reaction network rates. We illustrate via numerical examples how this novel technique could be useful to infer quantities pertaining to single-molecule dynamics (such as the distribution of the number of infections a single individual undergoes in a SIS model, or the time a single enzyme spends in the bound state) in a computationally efficient way.

Further, we consider the problem of comparing the dynamics of an original full reaction network with that of a collection of independent approximations of single-

molecule trajectories and provide explicit bounds on the error. Having the dynamics of the whole system approximated by a number of independent trajectories allows for computationally efficient simulation techniques, that are fully parallelizable. Moreover, since the DSA approximation is defined for all times, it does not suffer from the problem of exiting the state space as it is known to happen in other methods such as diffusion approximations or tau leaping [4, 5, 12, 18]. Finally, the independence of the single-molecule trajectories also allows for much simplified statistical inferential procedures. Such applications were already considered in the context of SIR networks in recent papers on the Covid pandemic [14, 23]. A thorough investigation of these techniques in general reaction networks is currently being conducted and will appear in a future work.

The paper is organized as follows: in Section 2 we provide the necessary concepts pertaining to reaction network theory followed by the result on the approximation in classical scaling in Section 3. In Section 4 we give a formal definition of single-molecule trajectories of what we refer to as species that are 'trackable'. In Section 5 we state our main results. In particular, in Section 5.1 we give the theorem on the Poisson process representation of the DSA approximation for a single-molecule trajectory, and give examples of its applications in Section 5.2. Finally, in Section 5.3 we state the result on the approximation of the original full network via independent single-molecule trajectories, and give numerical examples. Proofs and explicit error bounds are given in the Appendix A.

2. Background definitions.

2.1. Notation. We denote by $\mathbb{R}$, $\mathbb{R}_{>0}$, and $\mathbb{R}_{\geq 0}$ the real, positive real, and non-negative real numbers, respectively. Similarly, we denote by $\mathbb{Z}$, $\mathbb{Z}_{\geq 1}$, and $\mathbb{Z}_{\geq 0}$ the real, positive real, and non-negative real numbers, respectively. Given a number $r \in \mathbb{R}$, we denote by $|r|$ its absolute value, and by $\lfloor r \rfloor$ the largest $m \in \mathbb{Z}$ such that $m \leq r$.

Given a vectors $v \in \mathbb{R}^n$, we denote its ith component by v_i, for all $1 \leq i \leq n$. We further denote

$$\|v\|_\infty = \max_{1 \leq i \leq n} |v_i| \quad \text{and} \quad \lfloor v \rfloor = (\lfloor v_1 \rfloor, \ldots, \lfloor v_n \rfloor).$$

Given two vectors $u, v \in \mathbb{R}^n_{\geq 0}$, we write

$$u^v = \prod_{i=1}^{m} u_i^{v_i},$$

with the convention that $0^0 = 1$. We also write $u \geq v$ if the inequality holds component-wise. Furthermore, for any vector $v \in \mathbb{Z}^n_{\geq 0}$, we write

$$v! = \prod_{i=1}^{m} v_i!\,.$$

Given a set A, we denote its cardinality by $\#A$ or, if it leads to no ambiguity, by $|A|$. We assume the reader is familiar with basic notions from stochastic process theory, such as the definition of continuous-time Markov chains and Poisson processes [19].

Consider a sequence of random variables $\{X_n\}_{n \in \mathbb{Z}_{\geq 0}}$ and a random variable X, all defined on the same probability space and with values in a normed space $(E, \|\cdot\|)$.

We say that X_n converges in probability to X if for all $\eta \in \mathbb{R}_{>0}$

$$\lim_{n\to\infty} P\left(\|X_n - X\| > \eta\right) = 0.$$

Given a topological space E we will denote by $D_E[0,T]$ the set of right-continuous left-bounded functions defined from $[0,T]$ to E, endowed with the Skorohod J_1 topology. In particular, we say that the sequence of processes $\{X_n\}$ with sample paths in $D_E[0,T]$ converges in probability to the process X (or simply that X_n converges in probability to X) if the Skorohod distance between X_n and X converges to 0 in probability (for more details, see for example [11, Chapter 3]).

2.2. Stochastic reaction networks. A *reaction network* is a triple $\mathcal{G} = \{\mathcal{X}, \mathcal{C}, \mathcal{R}\}$, where (a) $\mathcal{X}$ is an ordered finite sequence of d symbols, called *species*; (b) $\mathcal{C}$ is a finite set of linear combinations of species over $\mathbb{Z}_{\geq 0}$, called *complexes*; (c) $\mathcal{R}$ is a finite set of elements of $\mathcal{C} \times \mathcal{C}$, called *reactions*. We assume that no element of the form (y, y) is in $\mathcal{R}$, for any complex y, even though our results do not depend on this assumption. Following the usual notation of reaction network Theory, we further denote a reaction $(y, y') \in \mathcal{R}$ by $y \to y'$. We finally assume that each complex appears in at least one reaction, and that each species has a positive coefficient in at least one complex. Under this assumption and up to ordering of the set of species, a reaction network is uniquely determined by the set $\mathcal{R}$, or equivalently by the directed graph $(\mathcal{C}, \mathcal{R})$, called *reaction graph*. As an example, consider the reaction graph

$$A + B \rightleftharpoons 2B, \quad B \longrightarrow C. \tag{2.1}$$

In this case, the associated species are A, B, and C, $\mathcal{C} = \{A + B, 2B, B, C\}$, and $\mathcal{R} = \{A + B \to 2B, 2B \to A + B, B \to C\}$.

In this paper we will implicitly identify $\mathbb{R}^{|\mathcal{X}|}$ with $\mathbb{R}^d$, and therefore each $S \in \mathcal{X}$ with a canonical basis vector of $\mathbb{R}^d$. With this in mind, the complexes are linear combination of species and can be therefore considered as vectors in $\mathbb{Z}^d_{\geq 0}$. As an example, if we order the species of (2.1) alphabetically, then the complex $A+B$ can be associated with the vector $(1, 1, 0)$, the complex $2B$ can be associated with $(0, 2, 0)$, the complex C with $(0, 0, 1)$, and so on. We will tacitly use the identification of complexes with integer vectors throughout the paper. Moreover, for each vector $v \in \mathbb{R}^d$ and for each species $S \in \mathcal{X}$ we denote by v_S the entry of v related to the canonical vector associated with S. We further define the *support* of v as $\mathrm{supp}(v) = \{S \in \mathcal{X} : v_S > 0\}$. As an example, with the species of (2.1) alphabetically ordered, the support of $(1, 1, 0)$ is $\{A, B\}$, the support of $(0, 2, 0)$ is $\{B\}$, and so on.

Deterministic and stochastic dynamical systems can be associated with a reaction network. The stochastic model is usually utilized when few individuals are present, so the stochastic component of the dynamic behaviour should not be ignored. In this case, the time evolution of the number of individuals of the different species is considered, for certain given propensities of the reactions to occur, and modeled via a continuous time Markov chain. More precisely, a *stochastic kinetics* for a reaction network $\mathcal{G}$ is a correspondence between a reaction $y \to y'$ and a *rate function* $\lambda_{y\to y'} \colon \mathbb{Z}^d_{\geq 0} \to \mathbb{R}_{\geq 0}$, such that $\lambda_{y\to y'}(x) > 0$ only if $x \geq y$. A *stochastic reaction system* is a continuous time Markov chain $\{X(t) : t \geq 0\}$ with state space $\mathbb{Z}^d_{\geq 0}$ and transition rates from a state x to a state x' defined by

$$q(x, x') = \sum_{\substack{y\to y'\in\mathcal{R} \\ y'-y=x'-x}} \lambda_{y\to y'}(x).$$

The associated generator is defined by

$$Af(x) = \sum_{y \to y' \in \mathcal{R}} \lambda_{y\to y'}(x)\Big(f(x+y'-y) - f(x)\Big)$$

for any function $f\colon \mathbb{Z}^d_{\geq 0} \to \mathbb{R}$ and any $x \in \mathbb{Z}^d_{\geq 0}$. Equivalently, the process X can be described by

$$X(t) = X(0) + \sum_{y \to y' \in \mathcal{R}} (y'-y) N_{y\to y'}\left(\int_0^\infty \lambda_{y\to y'}(X(s))ds\right),$$

where the processes $\{N_{y\to y'}\}_{y\to y'\in\mathcal{R}}$ are independent unit-rate Poisson processes. For more details on this representation, we refer to [6] or [11, Chapter 6].

In the deterministic setting, the concentration of the different species are assumed to evolve according to an ordinary differential equation (ODE). Specifically, a *deterministic kinetics* for a reaction network $\mathcal{G}$ is a correspondence between the reactions $y \to y'$ and the *rate function* $\lambda_{y\to y'}\colon \mathbb{R}^d_{\geq 0} \to \mathbb{R}_{\geq 0}$, such that $\lambda_{y\to y'}(x) > 0$ only if $x_i > 0$ whenever $y_i > 0$. A *deterministic reaction system* is the solution to the ordinary differential equation

$$\frac{d}{dt}Z(t) = \sum_{y\to y'\in\mathcal{R}} (y'-y)\lambda_{y\to y'}(x). \tag{2.2}$$

While our results hold in a more general scenario, all the simulations we show assume *mass-action kinetics*, a popular choice of kinetics derived by the assumption that all the reactants are well-mixed in the available volume [6]. Specifically, a stochastic reaction system is a *stochastic mass-action system* if for every reaction $y \to y' \in \mathcal{R}$ we have

$$\lambda_{y\to y'}(x) = \kappa_{y\to y'} \frac{x!}{(x-y)!}\mathbb{1}_{\{x\geq y\}},$$

for some positive constant $\kappa_{y\to y'}$ called *rate constant*. Similarly, a deterministic reaction system is a *deterministic mass-action system* if for every reaction $y \to y' \in \mathcal{R}$ we have

$$\lambda_{y\to y'}(x) = \kappa_{y\to y'} x^y,$$

for some positive constant $\kappa_{y\to y'}$ also called *rate constant*.

3. Classical scaling. Consider a reaction network $\mathcal{G} = \{\mathcal{X}, \mathcal{C}, \mathcal{R}\}$, and a family of stochastic kinetics $\{\lambda^V_{y\to y'} : y \to y' \in \mathcal{R}\}$ indexed by V. Let X^V denote the associated continuous time Markov chain. V should be thought to as a parameter expressing the volume, or the magnitude of the number of the present individuals. Under the following technical but reasonable assumption the classical scaling of [11,16] holds:

Assumption 3.1. We assume that for any reaction $y \to y' \in \mathcal{R}$ there exists a locally Lipschitz function $\lambda_{y\to y'}\colon \mathbb{R}^d_{\geq 0} \to \mathbb{R}^d_{\geq 0}$ such that for any compact set $K \subset \mathbb{R}^d_{\geq 0}$ we have

$$\lim_{V\to\infty} \sup_{z\in K} \left| \frac{\lambda^V_{y\to y'}(\lfloor Vz \rfloor)}{V} - \lambda_{y\to y'}(z)\right| = 0.$$

THEOREM 3.2. *Assume that Assumption 3.1 holds. Furthermore, assume that the random variables $X^V(0)/V$ converge in probability to a constant z^* as V goes to infinity. Finally, let $\{Z(t) : t \geq 0\}$ be the unique solution to (2.2) with $Z(0) = z^*$. Then, for any $\varepsilon > 0$ and any $T > 0$*

$$\lim_{V\to\infty} P\left(\sup_{t\in[0,T]} \left\| \frac{X^V(t)}{V} - Z(t) \right\|_\infty > \varepsilon\right) = 0.$$

Note that the distribution of the fate of a single molecule is not given, since the classical scaling concerns average dynamics. The goal of this paper is to address this issue, by providing a technique to simulate an approximation of the time evolution of a single observable species, as described in the next section.

4. Trackable species. We consider a special set of reactants, and assume that we can consistently follow the fate of a single molecule of these reactants through its different transformations, as for a single individual in the SI model. In general, the reactants whose dynamics we want to follow may be less than the chemical species listed in $\mathcal{X}$, or may be *portions* of them as in Example 4.5 below. To deal with this general setting, we introduce the set of *trackable species* as a set $\widetilde{\mathcal{X}}$ of symbols endowed with a function $\tau\colon \widetilde{\mathcal{X}} \to \mathcal{X} \cup \{0\}$. The different trackable species will identify the possible states of the molecules whose fate we want to follow. Every such state is taken by the reactant molecule when the molecule is (part of) one of the chemical species of $\mathcal{X}$. The function τ will link every trackable species with the corresponding species in $\mathcal{X}$. The number of trackable species defined in this way can be less than, equal to, or larger than the number of species. The set $\widetilde{\mathcal{X}}$ needs to include the special state Δ to denote the potential degradation of the tracked molecule, and we set $\tau\Delta = 0$. To simplify the notation, for all $x, y \in \mathbb{Z}^d_{\geq 0}$ and $\widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}$ we denote by $\theta_y(\widetilde{S}, x)$ the probability that a certain molecule of species $\tau(\widetilde{S})$ is chosen if $y_{\tau(\widetilde{S})}$ molecules are uniformly drawn out of $x_{\tau(\widetilde{S})}$ molecules of $\tau(\widetilde{S})$ available. Specifically,

$$\theta_y(\widetilde{S}, x) = \begin{cases} \frac{\binom{x_{\tau(\widetilde{S})}}{y_{\tau(\widetilde{S})}}^{-1}}{\binom{x}{y_{\tau(\widetilde{S})}}^{-1}} = \frac{y_{\tau(\widetilde{S})}}{x_{\tau(\widetilde{S})}} & \text{if } x_{\tau(\widetilde{S})} \geq y_{\tau(\widetilde{S})} \geq 1 \\ 0 & \text{otherwise} \end{cases}.$$

For completeness, we define $\theta_y(\Delta, x) = 0$. Finally, note that in reactions such as $2A \to B+C$ we can imagine a molecule of A is transformed into a molecule of B, while the other molecule of A turns into a molecule of C. If we are tracking the fate of A molecules and the reaction $2A \to B+C$ occurs, it is reasonable to assume the molecule we are tracking has a 50% change of turning into a molecule of B, and a 50% change of becoming a molecule of C. We denote these probabilities with $p_{2A\to B+C}(A, B)$ and $p_{2A\to B+C}(A, C)$, respectively, and in general allow for different value choices, as along as $p_{2A\to B+C}(A, B) + p_{2A\to B+C}(A, C) = 1$. The definition of *stochastic reaction system with trackable species* in the most general setting is below.

DEFINITION 4.1 (Stochastic reaction system with trackable species). *Let $\mathcal{G} = \{\mathcal{X}, \mathcal{C}, \mathcal{R}\}$ be a reaction network. Consider a family of stochastic kinetics $\{\lambda^V_{y\to y'} : y \to y' \in \mathcal{R}\}$ indexed by V, and let X^V denote the associated continuous time Markov chains. Let $\widetilde{\mathcal{X}}$ be a set of trackable species. We define the* stochastic reaction system with trackable species *as the continuous-time Markov chain (Y^V, X^V) with state space*

$\widetilde{\mathcal{X}} \times \mathbb{Z}^d_{\geq 0}$ *and transition rates*

$$q\Big((\Delta, x), (\widetilde{S}', x')\Big) = \mathbb{1}_{\{\widetilde{S}'\}}(\Delta) \sum_{\substack{y \to y' \in \mathcal{R} \\ y'-y = x'-x}} \lambda^V_{y \to y'}(x)$$

and for all $\widetilde{S} \neq \Delta$

$$q\Big((\widetilde{S}, x), (\widetilde{S}', x')\Big) = \sum_{\substack{y \to y' \in \mathcal{R} \\ y'-y = x'-x}} \Big((1 - \theta_y(\widetilde{S}, x))\mathbb{1}_{\{\widetilde{S}=\widetilde{S}'\}} + \theta_y(\widetilde{S}, x) p_{y \to y'}(\widetilde{S}, \widetilde{S}')\Big) \lambda^V_{y \to y'}(x),$$

where for all reactions $y \to y' \in \mathcal{R}$ *the following holds:*

- *for any* $\widetilde{S} \in \widetilde{\mathcal{X}}, \widetilde{S}' \in \widetilde{\mathcal{X}} \cup \{\Delta\}$ *we have* $0 \leq p_{y \to y'}(\widetilde{S}, \widetilde{S}') \leq 1$*;*
- $p_{y \to y'}(\widetilde{S}, \widetilde{S}') = 0$ *whenever* $\tau(\widetilde{S}) \notin \mathrm{supp}(y)$ *or* $\tau(\widetilde{S}') \notin \mathrm{supp}(y') \cup \Delta$*;*
- *if* $\tau(\widetilde{S}) \in \mathrm{supp}(y)$ *then*

$$\sum_{\widetilde{S}' \in \widetilde{\mathcal{X}} \,:\, \tau(\widetilde{S}') \in \mathrm{supp}(y') \cup \Delta} p_{y \to y'}(\widetilde{S}, \widetilde{S}') = 1.$$

In the above definition, the usual stochastic reaction system is coupled with the fate of a single trackable molecule: a trackable molecule in state $\widetilde{S}$ can transform whenever a reaction $y \to y'$ occurs, with a probability given by $\theta_y(\widetilde{S}, \widetilde{S}')$. By definition, the quantity $\theta_y(\widetilde{S}, \widetilde{S}')$ denotes precisely the probability that the tracked molecule takes part in the reaction $y \to y'$, assuming that the reacting molecules are uniformly chosen among those present. If that happens, the new state of the tracked molecule is drawn according to the probability distribution $\{p_{y \to y'}(\widetilde{S}, \widetilde{S}')\}_{\widetilde{S}' \in \mathrm{supp}(y') \cup \Delta}$ (see Example 4.4 for a case where this distribution is non-trivial). If the tracked molecule is irreversibly degraded, its state becomes Δ and cannot be changed. In what follows, we will sometimes identify the state space of Y^V, given by $\widetilde{\mathcal{X}}$, with the canonical basis of $\mathbb{R}^{|\widetilde{\mathcal{X}}|}$, similarly to how complexes are implicitly identified with vectors in $\mathbb{R}^d$.

The only technical requirement to have trackable species is that in every reaction, every piece of the reactants on the left-hand side of the reaction either transforms into a piece of the reactants on the right-hand side, or is discarded. Mathematically, mapping the chemical species of the left-hand side of the reaction with those on the right-hand side is always possible by mapping to Δ potential species in excess on the left-hand side, so in principle the requirements of Definition 4.1 can always be satisfied with the choice $\widetilde{\mathcal{X}} = \mathcal{X}$ and τ being the identity. Moreover, if we consider the physical system modelled by the reaction network, it is always true that reactions either transform molecules or degrade them. Hence, even when considering the physical meaning of the model, trackable species can always be defined to track the fate of every molecule of particular interest. In this case however some care should be taken to reflect the real physical changes caused by the reactions, and the set $\widetilde{\mathcal{X}}$ may need to be different from $\mathcal{X}$, as in Example 4.5.

Remark 4.2. The generator of a stochastic reaction system with trackable species, as defined in Definition 4.1, is given by

$$\mathcal{A}f(\Delta, x) = \sum_{y \to y' \in \mathcal{R}} \lambda^V_{y \to y'}(x)\Big(f(\Delta, x + y' - y) - f(\Delta, x)\Big)$$

and for $\widetilde{S} \neq \Delta$

$$\mathcal{A}f(\widetilde{S},x) = \sum_{y\to y'\in\mathcal{R}} (1-\theta_y(\widetilde{S},x))\lambda^V_{y\to y'}(x)\Big(f(\widetilde{S},x+y'-y)-f(\widetilde{S},x)\Big)$$

$$+ \sum_{y\to y'\in\mathcal{R}} \sum_{\widetilde{S}'\in \mathrm{supp}(y')\cup\Delta} \theta_y(\widetilde{S},x)p_{y\to y'}(\widetilde{S},\widetilde{S}')\lambda^V_{y\to y'}(x)\Big(f(\widetilde{S}',x+y'-y)-f(\widetilde{S},x)\Big),$$

for all functions $f\colon (\widetilde{\mathcal{X}})\times\mathbb{Z}^d_{\geq 0}\to\mathbb{R}$.

Example 4.3. Consider the SI reaction network described in (1.1), which we repeat here for convenience:

$$S + I \longrightarrow 2I. \tag{4.1}$$

In this case, we are interested in describing the history of susceptible individuals who become infected. The set of trackable species is therefore $\widetilde{\mathcal{X}} = \{\widetilde{S},\widetilde{I}\}$ with $\tau(\widetilde{S}) = S$ and $\tau(\widetilde{I}) = I$. Furthermore, we choose the probabilities $p_{S+I\to 2I}(\widetilde{S},\widetilde{I}) = 1$ and $p_{S+I\to 2I}(\widetilde{I},\widetilde{I}) = 1$. Alternatively, one can simply consider $\widetilde{\mathcal{X}} = \{\widetilde{S}\}$, with the understanding that whenever a susceptible individual gets infected we consider it as irreversibly degraded, and its state becomes Δ. In this case, $p_{S+I\to 2I}(\widetilde{S},\Delta) = 1$.

The state of single individuals can be tracked also in the more complex model

$$S + I \longrightarrow 2I, \quad I \longrightarrow S. \tag{4.2}$$

Here, the set of trackable species is $\{\widetilde{S},\widetilde{I}\}$, with $\tau(\widetilde{S}) = S$ and $\tau(\widetilde{I}) = I$, and the transformation probabilities are $p_{S+I\to 2I}(\widetilde{S},\widetilde{I}) = 1$, $p_{S+I\to 2I}(\widetilde{I},\widetilde{I}) = 1$, $p_{I\to S}(\widetilde{I},\widetilde{S}) = 1$. Here, relevant questions on the fate of a single individual could concern, for example, the number of infections it undergoes in a given time, or after how long the nth infection occurs. We can even extend the model to include migrations, and obtain

$$S + I \longrightarrow 2I, \quad I \longrightarrow S, \quad 0 \rightleftharpoons S, \quad 0 \rightleftharpoons I. \tag{4.3}$$

In this case, it is natural to assume $p_{S\to 0}(\widetilde{S},\Delta) = 1$ and $p_{I\to 0}(\widetilde{I},\Delta) = 1$. Relevant questions could involve, for example, the average number of infection a susceptible individual undergoes before migrating.

Example 4.4. Consider the following reaction network, where a protein P promotes its own phosphorylation:

$$2P \longrightarrow P + P^*, \quad P^* \longrightarrow P, P \longrightarrow 0. \tag{4.4}$$

Here, we may assume we are interested in observing the dynamics of a molecule of protein P. Hence, the set of trackable species is $\{\widetilde{P},\widetilde{P^*}\}$ with $\tau(\widetilde{P}) = P$ and $\tau(\widetilde{P^*}) = P^*$. It is natural to assume that the two molecules of P involved in the reaction $2P \to P+P^*$ have the same probability of being phosphorylated or serving as the reaction catalyst. Hence, $p_{2P\to P+P^*}(\widetilde{P},\widetilde{P}) = p_{2P\to P+P^*}(\widetilde{P},\widetilde{P^*}) = 1/2$. The other transformation probabilities are given by $p_{P^*\to P}(\widetilde{P^*},\widetilde{P}) = 1$ and $p_{P\to 0}(\widetilde{P},\Delta) = 1$.

Example 4.5. Consider the following reaction network, depicting a Michaelis-Menten mechanism where the product protein and the enzyme can spontaneously transform into each other:

$$E + S \rightleftharpoons C \longrightarrow E + P, \quad P \rightleftharpoons E. \tag{4.5}$$

In particular, the complex C represents a molecule of substrate S and enzyme bound together. When the bound is broken, it is natural to assume that the molecule of enzyme is released while the molecule of substrate is either released or transformed into the product P. Suppose we want to keep track of the history of a molecule of substrate S. If we were dealing with a classic Michaelis-Menten kinetics, i.e. without the reactions $P \rightleftharpoons E$, then we could simply consider as trackable species S, C, and P (or, more formally, S, C, and P with $\tau(\widetilde{S}) = S$, $\tau(\widetilde{C}) = C$, and $\tau(\widetilde{P}) = P$). Indeed, a molecule of substrate can be either in free state ($\widetilde{S}$), bound with the enzyme ($\widetilde{C}$), or transformed into a product ($\widetilde{P}$). Since the reactions $P \rightleftharpoons E$ are present, if we want to keep track of the fate of a molecule of substrate S we need to take into account the fact that it can transform into an enzyme, so E becomes a possible state of the molecule (more formally, $\widetilde{E}$ with $\tau(\widetilde{E}) = E$). The problem now is that we need to differentiate between the portion of C that E and S get transformed into by the reaction $E + S \to C$: the portion of C that E gets transformed into will become e free enzyme again via the reaction $C \to E + P$, while the portion S gets transformed into will become a product. In order to formally express these dynamics, we consider as set of trackable species $\{\widetilde{E}, \widetilde{S}, \widetilde{P}, \widetilde{C}_E, \widetilde{C}_S\}$, where $\widetilde{C}_E$ denotes we are tracking a molecule of E bound in the complex C, while $\widetilde{C}_S$ denotes we are tracking a molecule of S bound in C. The function τ associates every trackable species with its physical type: $\tau(\widetilde{E}) = E$, $\tau(\widetilde{S}) = S$, $\tau(\widetilde{P}) = P$, $\tau(\widetilde{C}_E) = C$, and $\tau(\widetilde{C}_S) = C$. The transformation probabilities are given by

$$\begin{array}{lll} p_{E+S\to C}(\widetilde{E}, \widetilde{C}_E) = 1 & p_{C\to E+S}(\widetilde{C}_E, \widetilde{E}) = 1 & p_{C\to E+P}(\widetilde{C}_E, \widetilde{E}) = 1 \\ p_{E+S\to C}(\widetilde{S}, \widetilde{C}_S) = 1 & p_{C\to E+S}(\widetilde{C}_S, \widetilde{S}) = 1 & p_{C\to E+P}(\widetilde{C}_S, \widetilde{P}) = 1 \\ p_{P\to E}(\widetilde{P}, \widetilde{E}) = 1 & p_{E\to P}(\widetilde{E}, \widetilde{P}) = 1 & \end{array} \tag{354}$$

Remark 4.6. The interpretation of a stochastic reaction system with trackable species is that of a regular stochastic reaction system with the subsequent tranformations of a given particle being tracked. If the initial state $Y^V(0)$ of the tracked molecule is not present in the initial $X^V(0)$, that is if $X^V_{\tau(Y^V(0))}(0) = 0$, then the initial condition of (Y^V, X^V) is not consistent with the interpretation of the process. The process (Y^V, X^V) is still well-defined and its evolution can be studied, but its interpretation is no longer valid. In order to obtain meaningful results, we therefore tacitly assume that $X^V_{\tau(Y^V(0))}(0) > 0$, even if we do not require it formally.

4.1. Representation as a regular stochastic reaction network. In this section we show how a reaction network with trackable species (Y^V, X^V) can be realized as a regular stochastic reaction network with species set given by $\widetilde{\mathcal{X}} \sqcup \mathcal{X}$, where $\sqcup$ denotes a disjoint union. In particular, the state space is $\mathbb{Z}_{\geq 0}^{|\widetilde{\mathcal{X}}|} \times \mathbb{Z}_{\geq 0}^d$, where for convenience we consider the first coordinates to refer to $\widetilde{\mathcal{X}}$, and the rest to the species of the original process $\mathcal{X}$. We denote by $(\widetilde{x}, x)$ a generic state in $\mathbb{Z}_{\geq 0}^{|\widetilde{\mathcal{X}}|} \times \mathbb{Z}_{\geq 0}^d$. Consider the set of reactions $\mathcal{R} \cup \widetilde{\mathcal{R}}$ where

$$\widetilde{\mathcal{R}} = \{\widetilde{S} + y \to \widetilde{S}' + y' \,:\, y \to y' \in \mathcal{R}, \widetilde{S}, \widetilde{S}' \in \widetilde{\mathcal{X}} \text{ and } p_{y\to y'}(\widetilde{S}, \widetilde{S}') > 0\} \tag{370}$$

and endow them with the following reaction rates:

$$\lambda^V_{y\to y'}(\widetilde{x}, x) = \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}} \widetilde{x}_{\widetilde{S}}(1 - \theta_y(\widetilde{S}, x))\lambda^V_{y\to y'}(x) \tag{372}$$

$$\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{x}, x) = \widetilde{x}_{\widetilde{S}}\theta_y(\widetilde{S}, x)p_{y\to y'}(\widetilde{S}, \widetilde{S}')\lambda^V_{y\to y'}(x). \tag{373}$$

Note that the second component of the process has the same transitions as X^V, with exactly the same rates. Hence, we can safely denote the process associated with the above stochastic reaction network by $(\widetilde{Y}^V, X^V)$. Note that the quantity $\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}} \widetilde{x}_{\widetilde{S}}$ is conserved by all possible transitions. Hence, if we consider an initial condition $(\widetilde{Y}(0), X(0))$ with $\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}} \widetilde{Y}_{\widetilde{S}}(0) = 1$, then at any time point t exactly one entry of the vector $\widetilde{Y}(t)$ is 1, and the other entries are zero. It follows that there is a bijection between the possible values of $\widetilde{Y}$ and $\widetilde{\mathcal{X}}$, given by the function $\mathrm{supp}(\widetilde{Y}(t))$. In this case, by identifying trackable species with vectors of the canonical basis of $\mathbb{R}^{|\widetilde{\mathcal{X}}|}$ as already done in the paper for the species in $\mathcal{X}$, the transition rates can be equivalently written as

$$\lambda^V_{y\to y'}(\widetilde{x}, x) = \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}} \mathbb{1}_{\{\widetilde{S}\}}(\widetilde{x})(1-\theta_y(\widetilde{S}, x))\lambda^V_{y\to y'}(x)$$

$$\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{x}, x) = \mathbb{1}_{\{\widetilde{S}\}}(\widetilde{x})\theta_y(\widetilde{S}, x)p_{y\to y'}(\widetilde{S}, \widetilde{S}')\lambda^V_{y\to y'}(x),$$

Hence, if $\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}} \widetilde{Y}_{\widetilde{S}}(0) = 1$ then the transitions and the rates of (Y^V, X^V) and $(\widetilde{Y}^V, X^V)$ coincide, and (Y^V, X^V) can be therefore realized as a stochastic reaction network with an appropriate initial condition. In particular, we can write

(4.6)
$$X^V(t) = X^V(0) + \sum_{y\to y'\in\mathcal{R}} (y'-y)N_{y\to y'}\left(\int_0^t \lambda^V_{y\to y'}(X^V(s))ds\right)$$

(4.7)
$$Y^V(t) = Y^V(0) + \sum_{y+\widetilde{S}\to y'+\widetilde{S}'\in\widetilde{\mathcal{R}}} (\widetilde{S}'-\widetilde{S})N_{y+\widetilde{S}\to y'+\widetilde{S}'}\left(\int_0^t \lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(Y^V(s), X^V(s))ds\right)$$

where N_r for $r\in\mathcal{R}\cup\widetilde{\mathcal{R}}$ are independent unit-rate Poisson processes. Note that with the above writing, all the processes in the set $\{(Y^V, X^V)\}_{V\in\mathbb{Z}_{\geq 1}}$ can be defined on the same probability space.

5. Results. In this section we state the main results of the current paper and their applications.

5.1. Classical scaling for the fate of a single molecule. In this section we state a law of large number for the process Y^V. In order to do this, we consider a family of stochastic reaction systems with trackable species (Y^V, X^V), with V varying in the integer numbers greater than one. We then assume that Assumption 3.1 is satisfied for some locally Lipschitz functions $\lambda_{y\to y'}$, and denote by Z the solution to (2.2). Hence, we know by Theorem 3.2 that $V^{-1}X^V$ will converge to Z path-wise with the uniform convergence topology over compact intervals of time, for V going to infinity.

In this section we express (Y^V, X^V) by means of independent unit-rate Poisson processes, as in (4.6) and (4.7). With the notation introduced in the previous section in mind, we have the following first technical result:

LEMMA 5.1. *Assume that Assumption 3.1 holds. Then, for any $\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}$, any $w\in\widetilde{\mathcal{X}}$, and any compact set $K\subset\mathbb{R}^d_{>0}$ we have*

$$\lim_{V\to\infty}\sup_{z\in K}\left|\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}}(w, \lfloor Vz\rfloor) - \lambda_{y\to y'}(w, z)\right| = 0, \tag{5.1}$$

where the function $\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}\colon \widetilde{\mathcal{X}}\times\mathbb{R}^d_{\geq 0}$ *is defined as*

$$\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(w,z)=\mathbb{1}_{\{w\}}(\widetilde{S})p_{y\to y'}(\widetilde{S},\widetilde{S}')y_{\tau(\widetilde{S})}\frac{\lambda_{y\to y'}(z)}{z_{\tau(\widetilde{S})}}$$

if both $z_{\tau(\widetilde{S})}$ *and* $y_{\tau(\widetilde{S})}$ *are positive, and zero otherwise. Moreover, the function* $\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}$ *is locally Lipschitz if restricted to* $\widetilde{\mathcal{X}}\times\mathbb{R}^d_{>0}$.

Proof. If $y_{\tau(S)}=0$, then both $\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}}$ and $\lambda_{y\to y'}$ are constantly zero, hence (5.1) holds. If $y_{\tau(S)}$ is positive, then for all $z\in K$ we have

$$\left|\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}}(w,\lfloor Vz\rfloor)-\lambda_{y\to y'}(w,z)\right|=$$
$$\mathbb{1}_{\{w\}}(\widetilde{S})p_{y\to y'}(\widetilde{S},\widetilde{S}')\left|\theta_y(\widetilde{S},\lfloor Vz\rfloor)\lambda^V_{y\to y'}(\lfloor Vz\rfloor)-y_{\tau(S)}\frac{\lambda_{y\to y'}(z)}{z_{\tau(S)}}\right|$$

Let $m=\min_{z\in K}z_{\tau\widetilde{S}}$, which is positive because K is a compact set contained in $\mathbb{R}^d_{>0}$. If V is large enough such that $Vm>y_{\tau\widetilde{S}}$ then

$$\left|\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}}(w,\lfloor Vz\rfloor)-\lambda_{y\to y'}(w,z)\right|=$$
$$\mathbb{1}_{\{w\}}(\widetilde{S})p_{y\to y'}(\widetilde{S},\widetilde{S}')y_{\tau(S)}\left|\frac{\lambda^V_{y\to y'}(\lfloor Vz\rfloor)}{V\cdot(\lfloor Vz_{\tau(\widetilde{S})}\rfloor/V)}-\frac{\lambda_{y\to y'}(z)}{z_{\tau(S)}}\right|$$

Hence, (5.1) follows from Assumption 3.1 and

$$\max_{z\in K}\left|\frac{\lfloor Vz_{\tau(\widetilde{S})}\rfloor}{V}-z_{\tau(\widetilde{S})}\right|\leq\frac{1}{V}.$$

To conclude the proof, we only need to show that $\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}$ restricted to $\widetilde{\mathcal{X}}\times\mathbb{R}^d_{>0}$ is locally Lipschitz. However, this follows from it being the product (up to multiplication by a constant) of the two locally Lipschitz functions $z\mapsto 1/z_{\tau(\widetilde{S})}$ and $\lambda_{y\to y'}$. ∎

The main goal of this section is to prove a classical scaling limit for a single-molecule trajectory. To this aim, define the process Y by

(5.2)
$$Y(t)=Y(0)+\sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}}(\widetilde{S}'-\widetilde{S})N_{\widetilde{S}+y\to\widetilde{S}'+y'}\left(\int_0^t\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(Y(s),Z(s))ds\right).$$

Then, the following result holds, where we implicitly identify the states of Y^V and Y with the canonical basis of $\mathbb{R}^{|\widetilde{\mathcal{X}}|}$. Note that the assumption that all the components of the solution Z are strictly positive in the time interval $[0,T]$ is made, but this is only a mild restriction to avoid unnecessary technicality, and is always verified under mass-action kinetics as long as $Z(0)\in\mathbb{R}^d_{>0}$ (see Remark 5.3). The proof of the result is postponed to Appendix A, where more precise bounds are given.

THEOREM 5.2. *Assume that Assumption 3.1 holds. Furthermore, assume that the random variables* $X^V(0)/V$ *converge in probability to some* $z^*\in\mathbb{R}^d_{>0}$ *as* V *goes to infinity, and let* $Z(0)=z^*$. *Assume that the solution* Z *to* (2.2) *with* $Z(0)=z^*$

exists over the interval $[0,T]$ *and that*

$$m = \min_{\substack{i=1,2,\dots,d \\ u\in[0,T]}} Z_i(u) > 0.$$

Finally, assume that $Y^V(0) = Y(0)$ *for all positive integers* V*. Then*

$$\lim_{V\to\infty} \sup_{t\in[0,T]} P\left(Y^V(t) \neq Y(t)\right) = \lim_{V\to\infty} \sup_{t\in[0,T]} E\left[\|Y^V(t) - Y(t)\|_\infty\right] = 0. \tag{5.3}$$

Remark 5.3. If we consider mass-action kinetics, then the deterministic solutions never touch the boundaries, provided that the initial condition is strictly positive [22]. In this case, the existence of m as assumed in Theorem 5.2 is then guaranteed by $z^* \in \mathbb{R}^d_{>0}$.

Remark 5.4. Theorem 5.2 implies finite dimensional distribution convergence of Y^V to Y in the following sense: for all $0 \leq t_1 < t_2 < \cdots < t_n \leq T$ we have

$$P\left(\max_{1\leq i\leq n} \|Y^V(t_i) - Y(t_i)\|_\infty > 0\right) \leq \sum_{i=1}^{n} P\left(\|Y^V(t_i) - Y(t_i)\|_\infty > 0\right),$$

and the latter tends to 0 as V tends to ∞, under the conditions of Theorem 5.2.

Some simulations of the process Y are proposed in Figure 2 for the case of the SIS model (4.2). We conclude this section with the following result, concerning the convergence of Y^V to Y as processes with sample paths in $D_{\widetilde{\mathcal{X}}}[0,T]$. We note how this result is necessary for the convergence of continuous functionals of $D_{\widetilde{\mathcal{X}}}[0,T]$, as highlighted in Section 5.2.

THEOREM 5.5. *Assume that Assumption 3.1 holds. Furthermore, assume that the random variables* $X^V(0)/V$ *converge weakly to a constant* z^* *as* V *goes to infinity, and let* $Z(0) = z^*$*. Assume that the solution* Z *to* (2.2) *with* $Z(0) = z^*$ *exists over the interval* $[0,T]$ *and that*

$$m = \min_{\substack{S\in\mathcal{X} \\ u\in[0,T]}} Z_S(u) > 0.$$

Finally, assume that $Y^V(0) = Y(0)$ *for all positive integers* V*. Then* Y^V *converges in probability to* Y *as processes with sample paths in* $D_{\widetilde{\mathcal{X}}}[0,T]$ *(where we identify* $\widetilde{\mathcal{X}}$ *with the elements of the canonical basis of* $\mathbb{R}^{|\widetilde{\mathcal{X}}|}$ *and embed it with the metric* $\|\cdot\|_\infty$*, or any equivalent one).*

The proof is given in Appendix A.

5.2. Applications of Theorem 5.5. The convergence of Theorem 5.5 allows us to state convergence in probability of $f(Y^V)$ to $f(Y)$, where $f\colon D_{\widetilde{\mathcal{X}}}[0,T] \to \mathbb{R}$ is a functional that is continuous with respect to the Skorohod J_1 topology. Classical examples are $f(x) = \sup_{t\in[0,T]} \|x(t)\|_\infty$, $f(x) = \int_0^T \phi(x(s))ds$ for some continuous function ϕ, or $f(x) = \sup_{t\in[0,T]}(x(t) - x(t-))$ where $x(t-) = \lim_{h\uparrow t} x(h)$ (see for example [11, Chapter 3]). More concretely, a functional we may want to consider is the number of times an individual gets infected in the interval $[0,T]$, assuming the model of equation (4.2) is in place. We denote this functional by ψ. Note that the convergence of X^V/V to its deterministic fluid limit, as stated in Theorem 3.2, does not give any mean of inferring the distribution of $\psi(Y^V)$. However, knowing that

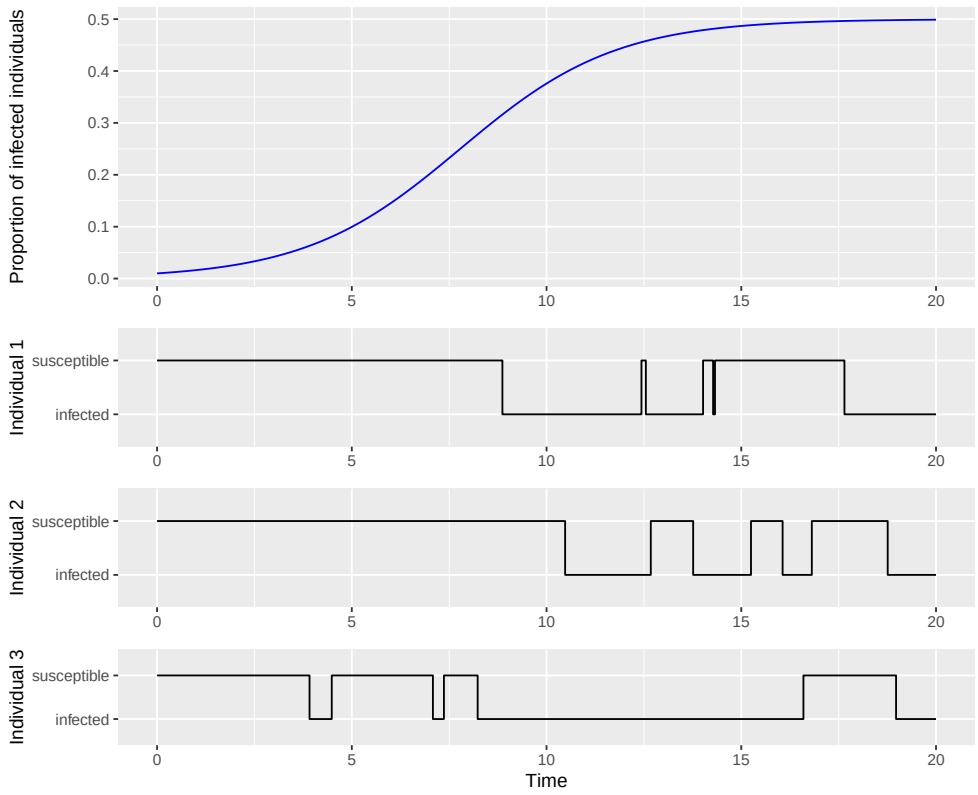

FIG. 2. ***The process Y in SIS model.*** *Consider the model (4.2), and let Y be as in (5.2). The first panel shows the concentration of infected individuals Z_I according to the deterministic solution to (2.2) with $Z_S(0) = 0.99$ and $Z_I(0) = 0.01$. Mass-action kinetics is assumed, with the rate constants of $S + I \to 2I$ and $I \to S$ being 1 and 0.5, respectively. According to (5.2), Z_I determines the rate at which the single-individual process Y turns from 'susceptible' to 'infected'. The last three panels show independent realizations of Y. The times in the x-axes of the four panels are aligned.*

$\psi(Y^V)$ converges in probability to $\psi(Y)$, if V is large enough we can approximate the distribution of the former by the distribution of the latter. Obtaining an estimate of the distribution of $\psi(Y)$ only requires the simulation of enough independent copies of Y, whose jump rates are deterministic and therefore do not require a simulation of X^V to be computed, as opposed to the much more expensive strategy of simulating multiple independent trajectories of (Y^V, X^V) via the Gillespie algorithm (which is especially cumbersome for large values of V). The empirical distributions obtained with he two strategies are compared in Figure 3. Similarly, we can apply our results to a Michaelis-Menten mechanism. Consider the model

$$E + S \rightleftharpoons C \longrightarrow E + P, \quad P \longrightarrow S, \tag{5.4}$$

where the enzyme activities counterbalances a spontaneous transformation of molecules of type P into molecules of type S. To measure the activity level of the enzymes, we may want to study for how long a randomly chosen enzyme molecule is in bound state C up to a given time T. Let us call this quantity $\upsilon(Y^V)$. The classical scal-

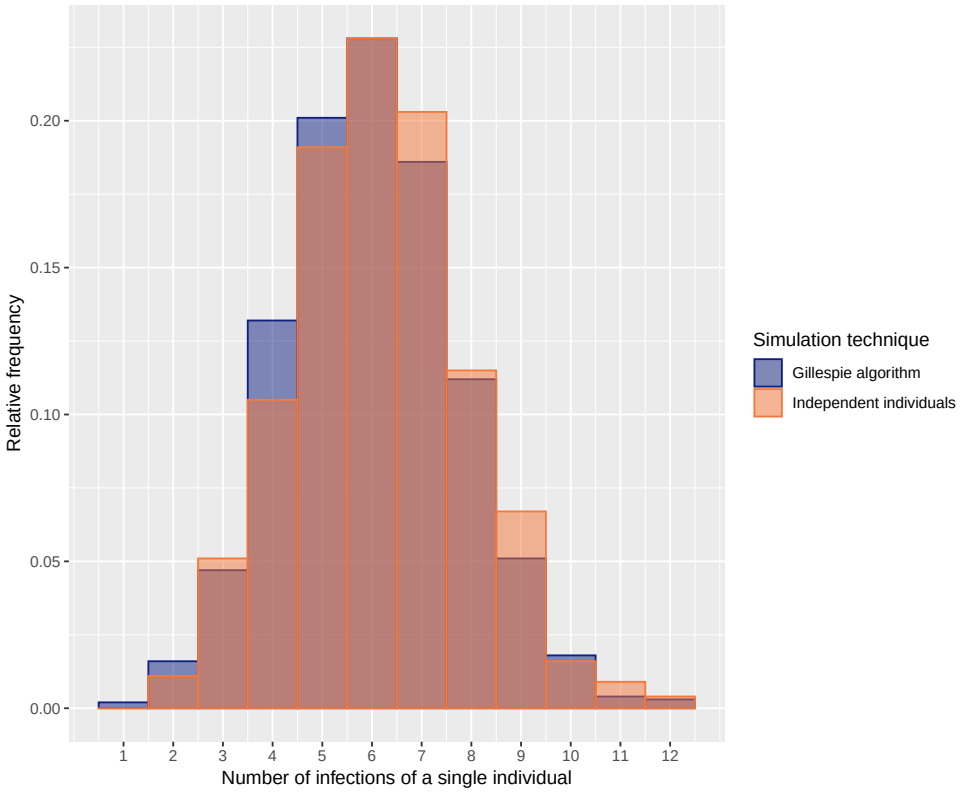

FIG. 3. ***Empirical distribution of number of infections in SIS model.*** *Consider the model (4.2), and let ψ be the number of infections a randomly selected individual undergoes up to time T. The empirical distributions of $\psi(Y^V)$ and $\psi(Y)$ are compared, the former obtained by the simulation of 1,000 independent copies of (Y^V, X^V) via the Gillespie algorithm (applied to the formulation in terms of usual stochastic reaction networks discussed in Section 4.1), and the latter obtained via the simulation of 1,000 copies of Y. Here, $V = 1,000$ and the initial portion of infected individuals is 1% (so we are initially close to the boundary and we may expect some minor discrepancy between X^V/V and its deterministic limit Z, see also Figure 5). Mass-action kinetics is assumed, with the rate constants of $S + I \to 2I$ and $I \to S$ being 1 and 0.5, respectively.*

ing of Theorem 3.2 does not allow for inference of the distribution of $\upsilon(Y^V)$, but Theorem 5.5 ensures that it converges to the distribution of $\upsilon(Y)$ as V tends to ∞. Figure 4 compares the empirical distributions of $\upsilon(Y^V)$ and $\upsilon(Y)$ obtained by the simulation of $1,000$ independent copies of (Y^V, X^V) and $1,000$ independent copies of Y, respectively. For this comparison we chose $V = 1,000$.

5.3. Approximating the system dynamics with single-molecule trajectories. Let $\overline{\mathcal{X}} \subseteq \mathcal{X}$ be the set of species that can be tracked in some form:

$$\overline{\mathcal{X}} = \{S \in \mathcal{X} \,:\, S = \tau(\widetilde{S}) \text{ for some } \widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}\}.$$

Moreover, let $\pi\colon \mathbb{R}^d \to \mathbb{R}^{|\overline{\mathcal{X}}|}$ be the projection of the state space onto the coordinates relative to the species in $\overline{\mathcal{X}}$. The aim of this section is to approximate the dynamics of $\pi(X^V)$ by means of a sum of *independent* processes distributed as in (5.2) (potentially with rescaled dynamics, as shown in the statement of Theorem 5.8). Note

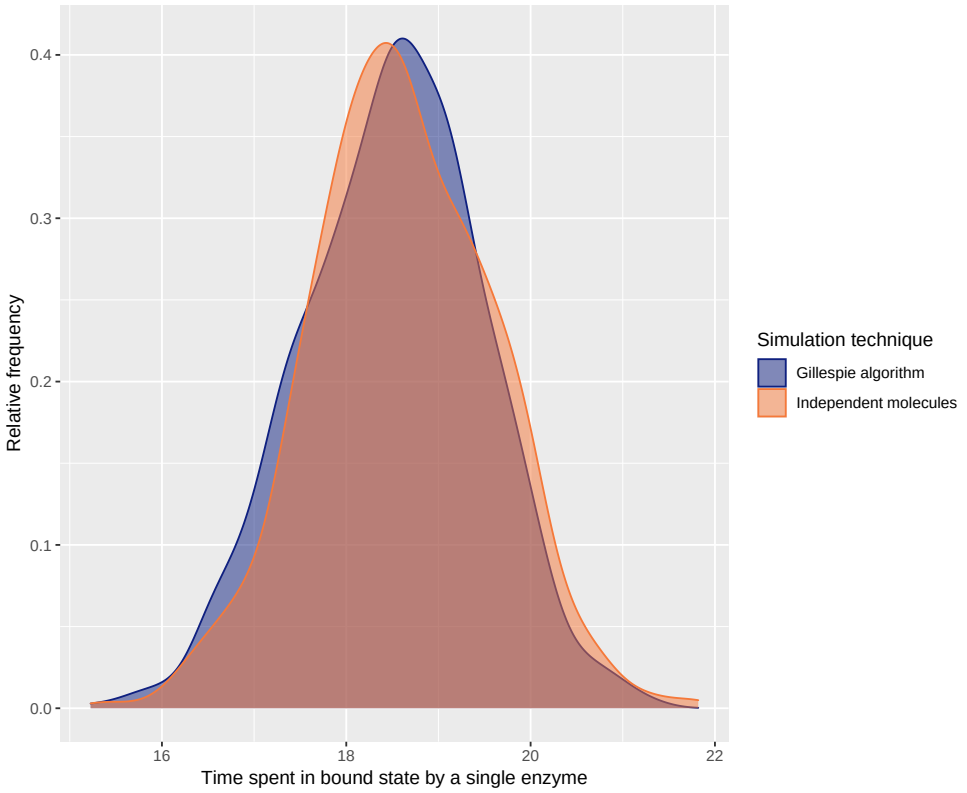

FIG. 4. ***Empirical density of time in bound state in Michaelis-Menten model.*** *Consider the model (5.4), and let v be the time a randomly selected molecule of enzyme is in bound state C up to time T. The empirical distributions of $v(Y^V)$ and $v(Y)$ are compared, the former obtained by the simulation of 1,000 independent copies of (Y^V, X^V) via the Gillespie algorithm (applied to the formulation in terms of usual stochastic reaction networks discussed in Section 4.1), and the latter obtained via the simulation of 1,000 copies of Y. Here, $V = 1,000$ and $Z(0) = X(0)/V = (0.5, 10, 0.5, 1)$, where the species are ordered as in E, S, C, P. Mass-action kinetics is assumed, with the rate constants of $E + S \to C$, $C \to E + S$, $C \to E + P$, and $P \to S$ being 1, 5, 1, and 0.5, respectively.*

that the goal of such an approximation is not to provide a faster simulation method than those present in the literature: our goal is to break down the dynamics of several correlated particles into a set of independent single-molecule trajectories which could be simulated simultaneously by a highly parallelizable algorithm. We begin by identifying each trackable species $\widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}$ with a different physical portions of the chemical species $\tau(\widetilde{S})$: m molecules of species $S \in \mathcal{X}$ are available at time t if and only if the quantity of each trackable species $\widetilde{S}$ satisfying $\tau(\widetilde{S}) = S$ is m at time t. Under this assumption, clearly the process X^V can be expressed in terms of the dynamics of its individual trackable species, which are typically not independent of each other. We further restrict ourselves to models that are *sub-conservative* with respect to the trackable species. This means that while trackable species can potentially be degraded (by entering the fictitious state Δ), their total mass never increases. Equivalently, we assume that each time a trackable species is created it is by transformation

of another trackable species. We assume sub-conservativeness because we want the single-molecule fates we track to be independent, while their agglomeration is still able to approximately describe the dynamics of the whole system. If we allowed for mass creation, we would need to introduce new molecules over time and track them. Defining the molecule creation times over a finite interval of time independently on each other is technically possible if the creation rate is deterministic: it is sufficient to first simulate a Poisson random variable counting the total number of new molecules in the finite time interval, then consider each creation time as independent on the others with probability density proportional to the deterministic creation rate. However, this procedure requires the introduction of further notation and for the sake of clarity we decided to only present the simpler case of sub-conservative models (with respect to the trackable species).

Assumption 5.6. Let (Y^V, X^V) be a family of stochastic reaction systems with trackable species. We assume that for each reaction $y \to y' \in \mathcal{R}$ and for each $\widetilde{S}' \in \widetilde{\mathcal{X}} \setminus \{\Delta\}$

$$\sum_{\widetilde{S},\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} p_{y\to y'}(\widetilde{S},\widetilde{S}') = y'_{\tau(\widetilde{S}')}$$

For all $S \in \overline{\mathcal{X}}, \widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}$ define

$$\tau^{-1}(S) = \{\widetilde{S}' \in \widetilde{\mathcal{X}} \,:\, \tau(\widetilde{S}') = S\} \quad \text{and} \quad \alpha(S) = \#\tau^{-1}(S)$$

The sub-conservation of the model with respect to the trackable species is formally stated as follows.

LEMMA 5.7. *Let (Y^V, X^V) be a family of stochastic reaction systems with trackable species satisfying Assumption 5.6. Then, for all $V \in \mathbb{Z}_{\geq 1}$ and for all $t \in \mathbb{R}_{>0}$*

$$\|\pi(X^V(t))\|_1 \leq \sum_{S\in\overline{\mathcal{X}}} \alpha(S) X^V_S(t) \leq \sum_{S\in\overline{\mathcal{X}}} \alpha(S) X^V_S(0). \tag{5.5}$$

Proof. The first inequality of (5.5) simply follows from the fact that the quantities $\alpha(S)$ are greater than or equal to 1. For the second inequality, simply note that if a reactions $y \to y' \in \mathcal{R}$ occurs at time t, then

$$\begin{aligned}
\sum_{S\in\overline{\mathcal{X}}} \alpha(S) X^V_S(t) - \sum_{S\in\overline{\mathcal{X}}} \alpha(S) X^V_S(t-) &= \sum_{S\in\overline{\mathcal{X}}} \alpha(S) y'_S - \sum_{S\in\overline{\mathcal{X}}} \alpha(S) y_S \\
&= \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y'_{\tau(\widetilde{S}')} - \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} \\
&= \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{\widetilde{S},\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} p_{y\to y'}(\widetilde{S},\widetilde{S}') - \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} \\
&\leq \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} - \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} = 0.
\end{aligned}$$

Note that in the third equality we used Assumption 5.6, and in the last equality we used $\sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} p_{y\to y'}(\widetilde{S},\widetilde{S}') \leq 1$. Since the quantity $\sum_{S\in\overline{\mathcal{X}}} \alpha(S) X^V_S$ is not increasing with the occurrence of a reaction, (5.5) is proven. ∎

The main result of this section is the following one, a more detailed version of which is proven in the Appendix. In particular, in Theorem A.5 a convergence rate of the order of $e^{-C\sqrt{V}}$ for a positive constant C is proven, provided that the initial conditions of X^V and $\widetilde{X}^V$ are close enough.

THEOREM 5.8. *Assume that Assumptions 3.1 and 5.6 are satisfied, and consider a family of stochastic reaction systems with trackable species (Y^V, X^V). Assume that $V^{-1}X^V(0)$ converges in distribution to some $z^* \in \mathbb{R}^d_{>0}$ as V goes to infinity and $E[\pi(X^V(0))] < \infty$ for all $V \in \mathbb{Z}_{\geq 1}$. Let $\widetilde{X}^V(0) = \lfloor Vz^* \rfloor$ and define the process $\widetilde{X}^V$ by*

$$\widetilde{X}^V(t) = \sum_{\widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}} \sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)} \frac{\tau(Y^{\widetilde{S},i}(t))}{\alpha(\tau(Y^{\widetilde{S},i}(t)))}, \tag{5.6}$$

where the processes $(Y^{\widetilde{S},i})_{\widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}, i \in \mathbb{Z}_{\geq 1}}$ are independent and satisfy

$$Y^{\widetilde{S},i}(t) = \widetilde{S} + \sum_{\widetilde{S}' + y \to \widetilde{S}'' + y' \in \widetilde{\mathcal{R}}} (\widetilde{S}'' - \widetilde{S}') N^{\widetilde{S},i}_{\widetilde{S}' + y \to \widetilde{S}'' + y'} \left(\int_0^t \lambda_{\widetilde{S}' + y \to \widetilde{S}'' + y'}(Y^{\widetilde{S},i}(u), Z(u)) du \right),$$

for a family of independent, identically distributed unit-rate Poisson processes $\{N^{\widetilde{S},i}_r\}_{\widetilde{S} \in \widetilde{\mathcal{X}} \setminus \{\Delta\}, i \in \mathbb{Z}_{\geq 1}, r \in \widetilde{\mathcal{R}}}$. Then,

$$\lim_{V \to \infty} E\left[\sup_{0 \leq s \leq t} \left\| \frac{\pi(X^V(t))}{V} - \frac{\widetilde{X}^V(t)}{V} \right\| \right] = 0.$$

Note that in the definition of $\widetilde{X}^V$ above we consider as many independent single-molecule trajectories as many trackable molecules are in the system at time 0. A natural question is whether a good approximation of the original model X^V can be obtained by considering the agglomeration of less independent single-molecule trajectories. However, a detailed study of the error in this case is out of the scope of the present paper.

Example 5.9. Consider the SIS model of equation (4.2). We assume $X^V_S(0) = 0.99V$ and $X^V_I(0) = 0.01V$, and let $V = 1,000$. We wish to approximate the number of susceptible individuals by

$$\frac{X^V_S(t)}{V} \approx \frac{\widetilde{X}^V_S(t)}{V}.$$

In order to test the performance of the above approximation, we simulate 100 independent copies of X^V and $\widetilde{X}$, and plot them against each other in Figure 5. It is perhaps not surprising to note a higher variance for the trajectories of X^V with respect of those of $\widetilde{X}^V$: the former is the result of several single-molecule trajectories that are naturally correlated with each other, specifically the rate at which a single molecule changes state is stochastic and given by the current state of all the other molecules. In the approximation, the dynamics of the single tracked molecules are independent and their rate of change from one state to another state are purely deterministic, which leads to fewer stochastic fluctuations. However, we do observe a discrepancy between the two models only at the beginning of the trajectories, when

the number of infected individuals is rather low (only 10 individuals in the initial condition) and the deterministic approximation given by Theorem 3.2 is perhaps not yet accurate enough. As a matter of fact, Figure 6 shows that the difference in variance is considerably reduced if the initial counts of infected individuals is increased to 100.

We are interested in bounding

$$P\left(\sup_{0\le t\le T}\left|\frac{X_S^V(t)}{V}-\frac{\widetilde{X}_S^V(t)}{V}\right|>\varepsilon\right), \tag{5.7}$$

for a fixed $\varepsilon \in \mathbb{R}_{>0}$. Assume mass-action kinetics and let κ_1 and κ_2 be the rate constants of $S+I\to 2I$ and $I\to S$, respectively. Moreover, assume for simplicity that $X^V(0)=\widetilde{X}^V(0)=VZ(0)$ and $X_S^V(0)+X_I^V(0)=V$. Since the total number of individual is conserved, for all $0\le t\le T$ we have $X_S^V(t)+X_I^V(t)=V$. By superposition there exist two independent unit-rate Poisson processes $\widetilde{N}_{S+I\to 2I}$ and $\widetilde{N}_{I\to S}$ such that for all $0\le t\le T$ and for a fixed V we have (with a simplified notation that does not take into account the initial values of the independent single individual

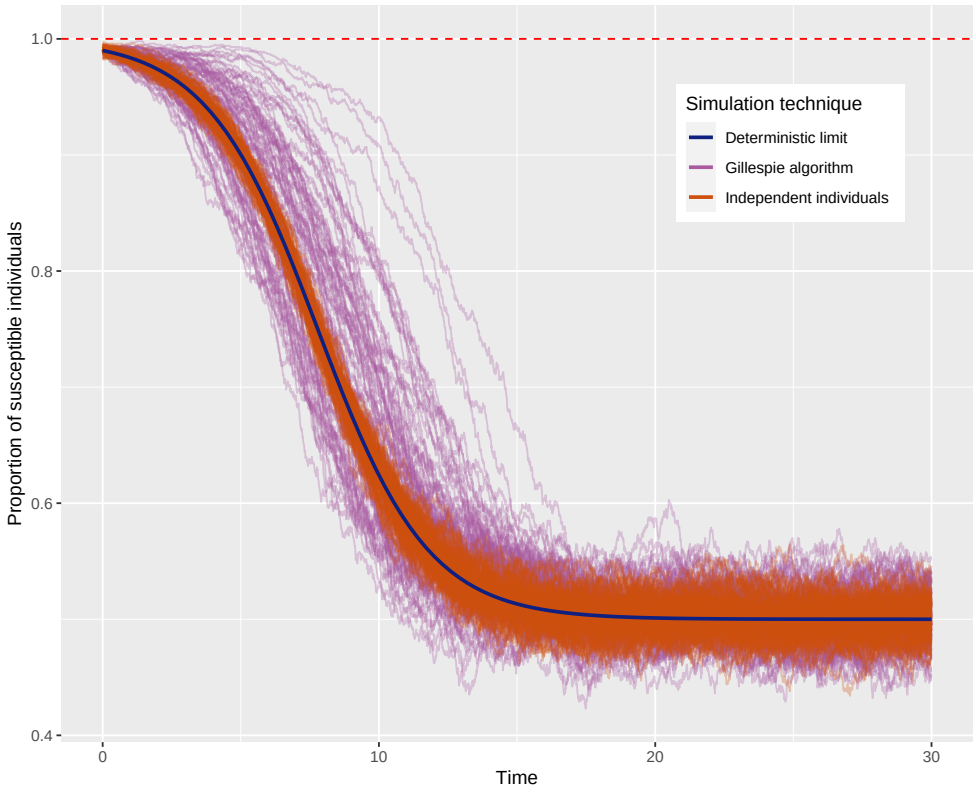

FIG. 5. ***Comparison in SIS model.*** *Comparison of 100 independent trajectories of* X_S^V/V *and* $\widetilde{X}_S^V/V$*, considering the SIS model described in* (4.2)*. Here,* $X_S^V(0)=0.99V$*,* $X_I^V(0)=0.01V$*, and* $V=1,000$*. Mass-action kinetics is assumed, with the rate constants of* $S+I\to 2I$ *and* $I\to S$ *being 1 and* 0.5*, respectively.*

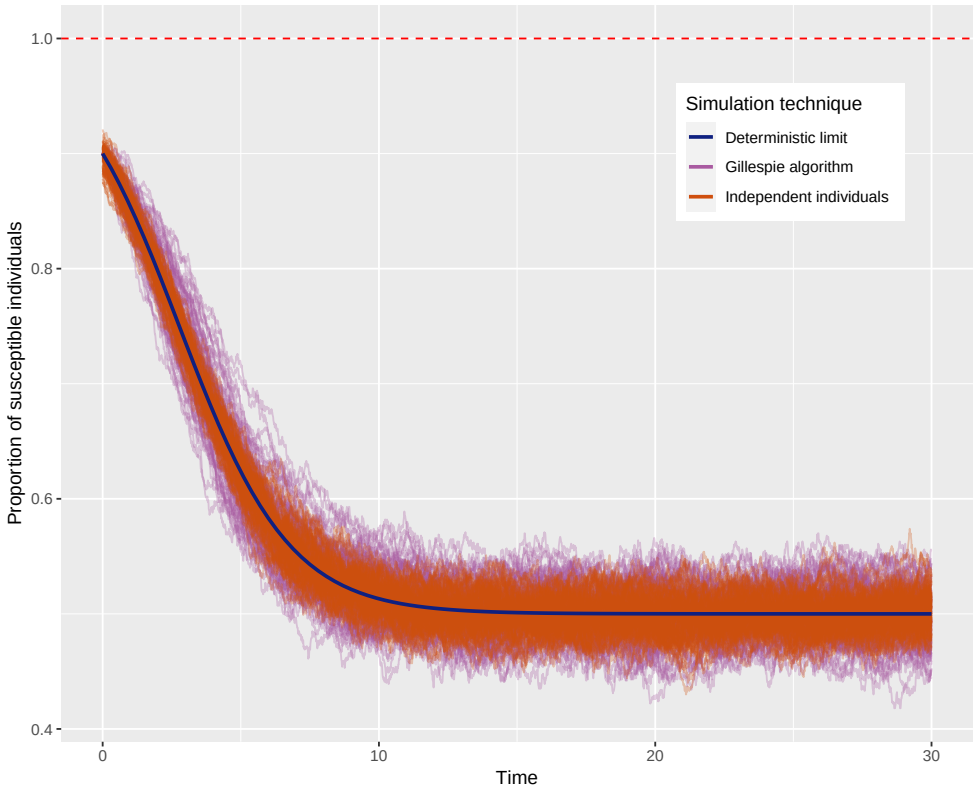

FIG. 6. ***Comparison in SIS model.*** *Comparison of 100 independent trajectories of* X_S^V/V *and* $\widetilde{X}_S^V/V$*, considering the SIS model described in* (4.2)*. Here,* $X_S^V(0) = 0.9V$*,* $X_I^V(0) = 0.1V$*, and* $V = 1,000$*. Mass-action kinetics is assumed, with the rate constants of* $S + I \to 2I$ *and* $I \to S$ *being 1 and* 0.5*, respectively.*

trajectories)

$$\widetilde{N}_{S+I\to 2I}\left(\int_0^t \kappa_1 \widetilde{X}_S^V(u) Z_I(u) du\right) = \sum_{i=1}^{V} N^i_{\widetilde{S}+S+I\to\widetilde{I}+2I}\left(\int_0^t \mathbb{1}_{\{\widetilde{S}\}}(Y^i(u)) Z_I(u) du\right)$$

$$\widetilde{N}_{I\to S}\left(\int_0^t \kappa_2 \widetilde{X}_I^V(u) du\right) = \sum_{i=1}^{V} N^i_{\widetilde{I}+I\to\widetilde{S}+S}\left(\int_0^t \mathbb{1}_{\{\widetilde{I}\}}(Y^i(u)) du\right).$$

Then,

$$\left|\frac{X_S^V(t)}{V} - \frac{\widetilde{X}_S^V(t)}{V}\right| \le \Delta(t) + \frac{1}{V}\int_0^t \kappa_1 X_S^V(u)\left|\frac{X_I^V(u)}{V} - Z_I(u)\right| du$$
$$+ \int_0^t \kappa_1 \left|\frac{X_S^V(u)}{V} - \frac{\widetilde{X}_S^V(u)}{V}\right| Z_I(u) du + \int_0^t \kappa_2 \left|\frac{X_I^V(u)}{V} - \frac{\widetilde{X}_I^V(u)}{V}\right| du,$$

where

$$\begin{aligned}\Delta(t) =& \frac{1}{V}\left|N_{S+I\to 2I}\left(\int_0^t \frac{\kappa_1}{V}X_S^V(u)X_I^V(u)du\right) - \int_0^t \frac{\kappa_1}{V}X_S^V(u)X_I^V(u)du\right| \\ &+ \frac{1}{V}\left|N_{I\to S}\left(\int_0^t \kappa_2 X_I^V(u)du\right) - \int_0^t \kappa_2 X_I^V(u)du\right| \\ &+ \frac{1}{V}\left|\widetilde{N}_{S+I\to 2I}\left(\int_0^t \kappa_1 \widetilde{X}_S^V(u)Z_I(u)du\right) - \int_0^t \kappa_1 \widetilde{X}_S^V(u)Z_I(u)du\right| \\ &+ \frac{1}{V}\left|\widetilde{N}_{I\to S}\left(\int_0^t \kappa_2 \widetilde{X}_I^V(u)du\right) - \int_0^t \kappa_2 \widetilde{X}_I^V(u)du\right|.\end{aligned}$$

Using $X_I^V(t) = V - X_I^V(t)$ and $Z_I(t) \le 1$ for all $0 \le t \le T$ we obtain

$$\begin{aligned}\left|\frac{X_S^V(t)}{V} - \frac{\widetilde{X}_S^V(t)}{V}\right| \le \Delta(t) &+ \int_0^t \kappa_1 \left|\frac{X_I^V(u)}{V} - Z_I(u)\right| du \\ &+ \int_0^t (\kappa_1+\kappa_2)\left|\frac{X_S^V(u)}{V} - \frac{\widetilde{X}_S^V(u)}{V}\right| du.\end{aligned}$$

By taking the supremum on $0 \le t \le T$ on both sides and by applying the Gronwall inequality, we have

$$\sup_{0\le t\le T}\left|\frac{X_S^V(t)}{V} - \frac{\widetilde{X}_S^V(t)}{V}\right| \le \left(\sup_{0\le t\le T}\Delta(t) + \kappa_1 T \sup_{0\le t\le T}\left|\frac{X_I^V(u)}{V} - Z_I(u)\right|\right) e^{(\kappa_1+\kappa_2)T}.$$

For notational convenience, let $\nu = \varepsilon e^{-(\kappa_1+\kappa_2)T}$. Hence, (5.7) is smaller than

$$P\left(\sup_{0\le t\le T}\Delta(t) > \frac{\nu}{2}\right) + P\left(\sup_{0\le t\le T}\left|\frac{X_I^V(u)}{V} - Z_I(u)\right| > \frac{\nu}{2\kappa_1 T}\right). \tag{5.8}$$

By noting that $P(\sup_{0\le t\le T}\Delta(t) > \nu/2)$ is smaller than

$$\begin{aligned}&P\left(\sup_{0\le t\le T}\frac{1}{V}\left|N_{S+I\to 2I}\left(\int_0^t \frac{\kappa_1}{V}X_S^V(u)X_I^V(u)du\right) - \int_0^t \frac{\kappa_1}{V}X_S^V(u)X_I^V(u)du\right| > \frac{\nu}{8}\right) \\ &+ P\left(\sup_{0\le t\le T}\frac{1}{V}\left|N_{I\to S}\left(\int_0^t \kappa_2 X_I^V(u)du\right) - \int_0^t \kappa_2 X_I^V(u)du\right| > \frac{\nu}{8}\right) \\ &+ P\left(\sup_{0\le t\le T}\frac{1}{V}\left|\widetilde{N}_{S+I\to 2I}\left(\int_0^t \kappa_1 \widetilde{X}_S^V(u)Z_I(u)du\right) - \int_0^t \kappa_1 \widetilde{X}_S^V(u)Z_I(u)du\right| > \frac{\nu}{8}\right) \\ &+ P\left(\sup_{0\le t\le T}\frac{1}{V}\left|\widetilde{N}_{I\to S}\left(\int_0^t \kappa_2 \widetilde{X}_I^V(u)du\right) - \int_0^t \kappa_2 \widetilde{X}_I^V(u)du\right| > \frac{\nu}{8}\right),\end{aligned}$$

we obtain that (5.8) is smaller than

$$\begin{aligned}&12\exp\left(\frac{\kappa_1 eT}{2} - \frac{\nu}{24}\sqrt{V}\right) + 12\exp\left(\frac{\kappa_2 eT}{2} - \frac{\nu}{24}\sqrt{V}\right) \\ &+ 6\exp\left(\frac{\kappa_1 eT}{2}\left(1+\frac{\nu}{\kappa_1 T}\right)^2 + \frac{\kappa_2 eT}{2}\left(1+\frac{\nu}{\kappa_1 T}\right) - \frac{\nu}{12\kappa_1 T}e^{-T(\kappa_1-\kappa_2)-\nu}\sqrt{V}\right)\end{aligned}$$

by Lemma A.1 and Theorem A.2 (for the special case of the SIS model, see Example A.3). We note that $\exp(h)$ is defined as e^h for all real numbers h. It follows that (5.7) tends to 0 as V tends to ∞ with the same rate as $e^{-C\sqrt{V}}$ for some positive constant C. This is always the case, and bounds for more general models are provided by Theorem A.5.

Appendix A. Proofs and explicit bounds.

In this section we give proofs for the results stated above, together with more precise bounds on the quantities of interest. To this aim, we first define the following quantities: for all $V \in \mathbb{Z}_{\geq 1}$ and $\varepsilon \in \mathbb{R}_{>0}$ let

$$\mathcal{A}_{V,\varepsilon,t} = \left\{ \sup_{u\in[0,t]} \left\| \frac{X^V(u)}{V} - Z(u) \right\|_\infty \leq \varepsilon \right\} \quad \text{and} \quad p^{V,\varepsilon,t} = P(\mathcal{A}^c_{V,\varepsilon,t}) = 1 - P(\mathcal{A}_{V,\varepsilon,t}),$$

where the superscript "c" denotes the complement. Note that, for any fixed for fixed V and ε, the sequence of events $\mathcal{A}_{V,\varepsilon,t}$ is monotone in t, and $p^{V,\varepsilon,t}$ is a non-decreasing function of t attaining its maximum for the value $t = T$.

Define the $\mathbb{Z}^d_{\geq 0}$-valued process $X^{V,\varepsilon}$ on $[0,T]$ in the following way: for any $S \in \mathcal{X}$ and any $t \in [0,T]$, let

$$X^{V,\varepsilon}_S(t) = \min\{\max\{X^V_S(t), VZ_S(t) - V\varepsilon\}, VZ_S(t) + V\varepsilon\}. \tag{A.1}$$

Hence, by definition for all $t \in \mathbb{R}_{>0}$

$$\left\| \frac{X^{V,\varepsilon}(t)}{V} - Z(t) \right\|_\infty \leq \varepsilon.$$

Moreover, define the process $\hat{X}^{V,\varepsilon}$ by

$$\hat{X}^{V,\varepsilon}(t) = X^V(0) + \sum_{y\to y'\in\mathcal{R}} (y'-y) N_{y\to y'} \left(\int_0^t \lambda^V_{y\to y'}(X^{V,\varepsilon}(u))du \right)$$

for all $t \in [0,T]$, where the processes $N_{y\to y'}$ are the same as in (4.6). Note that for any $t \in [0,T]$ we have $\mathbb{1}_{\mathcal{A}_{V,\varepsilon,t}} X^{V,\varepsilon}(t) = \mathbb{1}_{\mathcal{A}_{V,\varepsilon,t}} X^V(t) = \mathbb{1}_{\mathcal{A}_{V,\varepsilon,t}} \hat{X}^{V,\varepsilon}(t)$. In particular, it follows that

$$\begin{aligned} \sup_{0\leq u\leq t} \left\| \frac{X^{V,\varepsilon}(u)}{V} - Z(u) \right\|_\infty &\leq \mathbb{1}_{\mathcal{A}_{V,\varepsilon,t}} \sup_{0\leq u\leq t} \left\| \frac{\hat{X}^{V,\varepsilon}(u)}{V} - Z(u) \right\|_\infty + \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,t}} \varepsilon \\ &\leq \sup_{0\leq u\leq t} \left\| \frac{\hat{X}^{V,\varepsilon}(u)}{V} - Z(u) \right\|_\infty . \end{aligned} \tag{A.2}$$

For any $t \in [0,T]$ and any $\varepsilon \in \mathbb{R}_{>0}$ let

$$\Omega^{\varepsilon,t}_1 = \{Z(u) + h \,:\, u \in [0,t], h \in \mathbb{R}^d, \|h\|_\infty \leq \varepsilon\} \cap \mathbb{R}^d_{\geq 0}$$

be the (one-dimensional) neighbourhood of the solution Z on the interval $[0,t]$ with amplitude ε, intersected with the non-negative orthant. Note that for all $t \in [0,T]$ we have $X^{V,\varepsilon}(t)/V \in \Omega^{\varepsilon,V}_1$. Similarly, let

$$\Omega^{\varepsilon,t}_2 = \{(Z(u) + h, Z(u) + h') \,:\, u \in [0,t], h, h' \in \mathbb{R}^d, \|h\|_\infty \leq \varepsilon, \|h'\|_\infty \leq \varepsilon\} \cap \mathbb{R}^{2d}_{\geq 0}$$

be the two-dimensional neighbourhood of the Z restricted to $[0,t]$ with amplitude ε, intersected with the non-negative orthant.

To conclude, it is convenient to introduce in this section a notation for centered Poisson processes: given a Poisson process N, we denote by $\overline{N}$ the process defined by $\overline{N}(t) = N(t) - t$ for all $t \in \mathbb{R}_{\geq 0}$. In order to bound $p^{V,\varepsilon,t}$ from above and prove Theorem 5.8 we need the following results concerning centered Poisson processes. For completeness, we provide a proof as we were not able to find it in the literature, even if small variations of Lemma A.1 are well-known and obtained as an application of Doob's inequality or Kolmogorov's maximal inequality.

LEMMA A.1. *Let N be a Poisson process and let $T, \varepsilon \in \mathbb{R}_{>0}$. Then, for all $n \in \mathbb{Z}_{\geq 1}$*

$$P\left(\sup_{t\in[0,nT]}\left|\frac{\overline{N}(t)}{n}\right| > \varepsilon\right) \leq 6\exp\left(\frac{e}{2}T - \frac{\varepsilon\sqrt{n}}{3}\right).$$

Proof. For all $j \in \mathbb{Z}_{\geq 1}$ and all $h \in \mathbb{R}_{>0}$ define

$$\Xi_j^h = \bigcup_{i=0}^{2^j h}\left\{\frac{i}{2^j}\right\}. \tag{A.3}$$

Since $\overline{N}$ is almost surely right continuous, we have that for all $n \in \mathbb{Z}_{\geq 1}$ and all $T \in \mathbb{R}_{>0}$

$$\sup_{t\in[0,nT]}\left|\frac{\overline{N}(t)}{n}\right| = \lim_{j\to\infty}\max_{t\in\Xi_j^{nT}}\left|\frac{\overline{N}(t)}{n}\right|$$

almost surely. Since for all $j \in \mathbb{Z}_{\geq 1}$ we have $\Xi_j^{nT} \subset \Xi_{j+1}^{nT}$, by continuity of the probability measure we have

$$P\left(\sup_{t\in[0,nT]}\left|\frac{\overline{N}(t)}{n}\right| > \varepsilon\right) = \lim_{j\to\infty} P\left(\max_{t\in\Xi_j^{nT}}\left|\frac{\overline{N}(t)}{n}\right| > \varepsilon\right).$$

By Etemadi's inequality we have

$$P\left(\max_{t\in\Xi_j^{nT}}\left|\frac{\overline{N}(t)}{n}\right| > \varepsilon\right) \leq 3\max_{t\in\Xi_j^{nT}} P\left(\left|\frac{\overline{N}(t)}{n}\right| > \frac{\varepsilon}{3}\right).$$

Moreover, for any real $\beta \in (0,1)$ and any real $t \in (0, nT)$ we have

$$\begin{aligned}
P\left(\left|\frac{\overline{N}(t)}{n}\right| > \frac{\varepsilon}{3}\right) &\leq P\left(\frac{\overline{N}(t)}{n} > \frac{\varepsilon}{3}\right) + P\left(-\frac{\overline{N}(t)}{n} > \frac{\varepsilon}{3}\right)\\
&= P\left(e^{\frac{n^\beta \overline{N}(t)}{n}} > e^{\frac{n^\beta \varepsilon}{3}}\right) + P\left(e^{-\frac{n^\beta \overline{N}(t)}{n}} > e^{\frac{n^\beta \varepsilon}{3}}\right)\\
&\leq 2\exp\left(-\frac{n^\beta\varepsilon}{3}\right)\exp\left(t(e^{n^{\beta-1}} - 1 - n^{\beta-1})\right)\\
&\leq 2\exp\left(-\frac{n^\beta\varepsilon}{3}\right)\exp\left(nT\frac{n^{2\beta-2}}{2}e^{n^{\beta-1}}\right),\\
&\leq 2\exp\left(-\frac{n^\beta\varepsilon}{3}\right)\exp\left(nT\frac{n^{2\beta-2}}{2}e\right),
\end{aligned}$$

where the inequality in the third line follows from the Markov's inequality and the known form of the moment generating function of a Poisson random variable, which leads to $E[e^{n^{\beta-1}\overline{N}(t)}] = e^{-n^{\beta-1}t}e^{t(e^{n^{\beta-1}}-1)}$ and $E[e^{-n^{\beta-1}\overline{N}(t)}] = e^{n^{\beta-1}t}e^{t(e^{-n^{\beta-1}}-1)}$. Hence, for all $n \in \mathbb{Z}_{\geq 1}$ we have that both $E[e^{n^{\beta-1}\overline{N}(t)}]$ and $E[e^{-n^{\beta-1}\overline{N}(t)}]$ are less than or equal to $e^{t(e^{n^{\beta-1}}-1-n^{\beta-1})}$. The inequality in the forth line derives from the Taylor expansion of the exponential function. By choosing $\beta = 1/2$ we have

$$P\left(\left|\frac{\overline{N}(t)}{n}\right| > \frac{\varepsilon}{3}\right) \leq 2\exp\left(-\frac{\varepsilon\sqrt{n}}{3}\right)\exp\left(\frac{e}{2}T\right),$$

which completes the proof. ∎

A.1. Estimates for $p^{V,\varepsilon,t}$. Many papers have focused on quantifying the distance between the process X^V and its fluid limit Z. Among these, we list [1–3, 13, 17, 20] with no claim of completeness. Here we use Lemma A.1 to show the following upper bound on $p^{V,\varepsilon,t}$. While similar estimates are known in the reaction network community, we give a formal proof of the bound we propose as we could not find it in the literature. Before stating the result, we define the following quantities:

$$R = \max_{y\to y'\in\mathcal{R}} \|y' - y\|_\infty,$$

$$\Lambda_0^{\varepsilon,t} = \sup_{z\in\Omega_1^{\varepsilon,t}} \sum_{y\to y'\in\mathcal{R}} \lambda_{y\to y'}(z), \quad \Lambda_1^{\varepsilon,t} = \int_0^t \Lambda_0^{\varepsilon,u} du$$

$$L_0^{\varepsilon,t} = \sup_{\substack{(z,z')\in\Omega_2^{\varepsilon,t}\\ z\neq z'}} \sum_{y\to y'\in\mathcal{R}} \frac{|\lambda_{y\to y'}(z) - \lambda_{y\to y'}(z')|}{\|z - z'\|_\infty}, \quad L_1^{\varepsilon,t} = \int_0^t L_0^{\varepsilon,u} du$$

$$\delta_0^{V,\varepsilon,t} = \sup_{z\in\Omega_1^{\varepsilon,t}} \sum_{y\to y'\in\mathcal{R}} \left|\frac{\lambda^V_{y\to y'}(\lfloor Vz\rfloor)}{V} - \lambda_{y\to y'}(z)\right|, \quad \delta_1^{V,\varepsilon,t} = \int_0^t \delta_0^{V,\varepsilon,u} du$$

$$\eta^{V,\varepsilon,t}(\gamma) = e^{-L_1^{2\varepsilon,t}}\gamma\varepsilon - \delta_1^{V,2\varepsilon,t},$$

where in the last definition γ is any real number in $(0, 1]$. Note that $\Lambda_0^{\varepsilon,t}$ and $\delta_0^{V,\varepsilon,t}$ are finite for any $t \in [0, T]$, since the solution Z exists up to time T and the functions $\lambda_{y\to y'}$ are locally Lipschitz by Assumption 3.1. The local Lipschitzianity of the functions $\lambda_{y\to y'}$ also implies that $L_0^{\varepsilon,t}$ is finite for all $\varepsilon \in \mathbb{R}_{>0}$ and $t \in [0, T]$. It also follows from Assumption 3.1 that $\delta_0^{V,\varepsilon,t}$ tends to zero as V tends to infinity. Furthermore, note that for fixed $V \in \mathbb{Z}_{\geq 1}$ and $\varepsilon \in \mathbb{R}_{>0}$, the quantities $\Lambda_0^{\varepsilon,t}, L_0^{\varepsilon,t}$, and $\delta_0^{V,\varepsilon,t}$ are all non-decreasing functions of t. As a consequence, for all $t \in [0, T], \varepsilon \in \mathbb{R}_{>0}$, and $V \in \mathbb{Z}_{\geq 1}$ we have

$$\Lambda_1^{\varepsilon,t} \leq t\Lambda_0^{\varepsilon,t}, \quad L_1^{\varepsilon,t} \leq tL_0^{\varepsilon,t}, \quad \text{and} \quad \delta_1^{V,\varepsilon,t} \leq t\delta_0^{V,\varepsilon,t}.$$

It follows that for all $t \in [0, T], \varepsilon \in \mathbb{R}_{>0}$, and $\gamma \in (0, 1]$ the quantity $\eta^{V,\varepsilon,t}(\gamma)$ tends to the positive quantity $e^{-L_1^{2\varepsilon,t}}\gamma\varepsilon$ as V tends to infinity. We can now state the following theorem.

THEOREM A.2. *For any $\varepsilon, t \in \mathbb{R}_{>0}$, any $\gamma \in (0, 1]$, and any $V \in \mathbb{Z}_{\geq 1}$ large enough such that $\eta^{V,2\varepsilon,t}(\gamma) > 0$, we have*

$$p^{V,\varepsilon,t} \leq p^{V,(1-\gamma)\varepsilon e^{-L_1^{2\varepsilon,t}},0} + 6\exp\left(\frac{e}{2}\Lambda_1^{2\varepsilon,t} + \frac{e}{2}\delta_1^{V,2\varepsilon,t} - \frac{1}{3R}\eta^{V,\varepsilon,t}(\gamma)\sqrt{V}\right)$$

Proof. First, note that

$$p^{V,\varepsilon,t} = P\left(\sup_{u\in[0,t]}\left\|\frac{X^V(u)}{V} - Z(u)\right\|_\infty > \varepsilon\right) = P\left(\sup_{u\in[0,t]}\left\|\frac{X^{V,2\varepsilon}(u)}{V} - Z(u)\right\|_\infty > \varepsilon\right)$$
$$= P\left(\sup_{u\in[0,t]}\left\|\frac{\hat{X}^{V,2\varepsilon}(u)}{V} - Z(u)\right\|_\infty > \varepsilon\right).$$

Moreover, by superposition, for all $V \in \mathbb{Z}_{\geq 1}$ and all $\varepsilon \in \mathbb{R}_{>0}$ we can define a unit-rate Poisson process $U^{V,2\varepsilon}$ coupled with X^V in such a way that for all $t \in \mathbb{R}_{\geq 0}$

$$U^{V,2\varepsilon}\left(\sum_{y\to y'\in\mathcal{R}}\int_0^t \lambda^V_{y\to y'}(X^{V,2\varepsilon}(u))du\right) = \sum_{y\to y'\in\mathcal{R}} N_{y\to y'}\left(\int_0^t \lambda^V_{y\to y'}(X^{V,2\varepsilon}(u))du\right).$$

Hence, by using (2.2) we have

$$\left\|\frac{\hat{X}^{V,2\varepsilon}(u)}{V} - Z(u)\right\|_\infty \leq \left\|\frac{\hat{X}^{V,2\varepsilon}(0)}{V} - Z(0)\right\|_\infty + \frac{R}{V}\left|\sum_{y\to y'\in\mathcal{R}}\overline{N}_{y\to y'}\left(\int_0^u \lambda^V_{y\to y'}(X^{V,2\varepsilon}(w))dw\right)\right|$$
$$+ \int_0^u \left|\sum_{y\to y'\in\mathcal{R}}\left(\frac{\lambda^V_{y\to y'}(X^{V,2\varepsilon}(w))}{V} - \lambda_{y\to y'}\left(\frac{X^{V,2\varepsilon}(w)}{V}\right)\right)dw\right|$$
$$+ \int_0^u \left|\sum_{y\to y'}\left(\lambda_{y\to y'}\left(\frac{X^{V,2\varepsilon}(w)}{V}\right) - \lambda_{y\to y'}(Z(w))\right)dw\right|$$
$$\leq \left\|\frac{X^V(0)}{V} - Z(0)\right\|_\infty + \frac{R}{V}\left|\overline{U}^{V,2\varepsilon}\left(\sum_{y\to y'\in\mathcal{R}}\int_0^u \lambda^V_{y\to y'}(X^{V,2\varepsilon}(w))dw\right)\right|$$
$$+ \delta_1^{V,2\varepsilon,u} + \int_0^u L_0^{2\varepsilon,w}\left\|\frac{X^{V,2\varepsilon}(w)}{V} - Z(w)\right\|_\infty dw$$

By using (A.2), by taking the supremum over $[0,t]$ on both sides we obtain

$$\sup_{0\leq u\leq t}\left\|\frac{\hat{X}^{V,2\varepsilon}(u)}{V} - Z(u)\right\|_\infty \leq \left\|\frac{X^V(0)}{V} - Z(0)\right\|_\infty$$
$$+ \frac{R}{V}\sup_{0\leq u\leq t}\left|\overline{U}^{V,2\varepsilon}\left(\sum_{y\to y'\in\mathcal{R}}\int_0^u \lambda^V_{y\to y'}(X^{V,2\varepsilon}(w))dw\right)\right|$$
$$+ \delta_1^{V,2\varepsilon,t} + \int_0^t L_0^{2\varepsilon,u}\sup_{0\leq w\leq u}\left\|\frac{\hat{X}^{V,2\varepsilon}(w)}{V} - Z(w)\right\|_\infty du.$$

By Gronwall's inequality we get

$$\sup_{0\leq u\leq t}\left\|\frac{\hat{X}^{V,2\varepsilon}(t)}{V} - Z(t)\right\|_\infty \leq e^{L_1^{2\varepsilon,t}}\left\|\frac{X^V(0)}{V} - Z(0)\right\|_\infty$$
$$+ \frac{Re^{L_1^{2\varepsilon,t}}}{V}\sup_{0\leq u\leq t}\left|\overline{U}^{V,2\varepsilon}\left(\sum_{y\to y'\in\mathcal{R}}\int_0^u \lambda^V_{y\to y'}(X^{V,2\varepsilon}(w))dw\right)\right|$$
$$+ e^{L_1^{2\varepsilon,t}}\delta_1^{V,2\varepsilon,t}.$$

By noting that for all $t \in \mathbb{R}_{\geq 0}$

$$\sup_{z \in \Omega_1^{2\varepsilon,t}} \sum_{y \to y' \in \mathcal{R}} \frac{\lambda^V_{y \to y'}(\lfloor Vz \rfloor)}{V} \leq \Lambda_0^{2\varepsilon,t} + \delta_0^{V,2\varepsilon,t},$$

we get

$$\begin{aligned} p^{V,\varepsilon,t} \leq & P\left(e^{L_1^{2\varepsilon,t}} \left\| \frac{X^V(0)}{V} - Z(0) \right\|_\infty > (1-\gamma)\varepsilon \right) \\ & + P\left(Re^{L_1^{2\varepsilon,t}} \sup_{0 \leq u \leq V(\Lambda_1^{2\varepsilon,t} + \delta_1^{V,2\varepsilon,t})} \left| \frac{\overline{U}^{V,2\varepsilon}(u)}{V} \right| + e^{L_1^{2\varepsilon,t}} \delta_1^{V,2\varepsilon,t} > \gamma\varepsilon \right) \end{aligned}$$

for any γ in $(0,1]$. the proof is concluded by Lemma A.1. ∎

Example A.3. Consider the SIS reaction network described in (4.2). In this case, in accordance with the classical mass-action choice of kinetics we have

$$\lambda^V_{S+I \to 2I}(x) = \frac{1}{V}\kappa_1 x_S x_I \quad \text{and} \quad \lambda^V_{I \to S}(x) = \kappa_2 x_I$$

for some positive constants κ_1 and κ_2. Hence, Assumption 3.1 is satisfied with

$$\lambda_{S+I \to 2I}(z) = \kappa_1 z_S z_I \quad \text{and} \quad \lambda_{I \to S}(z) = \kappa_2 z_I.$$

The corresponding solution Z exists for all non-negative times t, for all initial conditions $Z(0) = z^*$. Moreover, note that the sum of infected and susceptible individuals is kept constant, hence for all $t \in \mathbb{R}_{>0}$ we have $Z_S(t) + Z_I(t) = z^*_S + z^*_I = \|z^*\|_1$. In this case we can obtain the following rough estimates

$$R = 2, \quad \Lambda_0^{\varepsilon,t} \leq (\|z^*\|_1 + \varepsilon)[\kappa_1(\|z^*\|_1 + \varepsilon) + \kappa_2], \quad L_0^{\varepsilon,t} \leq \kappa_1(\|z^*\|_1 + \varepsilon) + \kappa_2,$$
$$\delta_0^{V,\varepsilon,t} = 0, \quad \eta^{V,\varepsilon,t} \geq \varepsilon e^{-t\kappa_1(\|z^*\|_1 + 2\varepsilon) + t\kappa_2}.$$

If we assume $X^V(0) = Vz^*$, then $p^{V,0,0} = 0$. It follows from Theorem A.2 with the choice $\gamma = 1$ that in this case

$$p^{V,\varepsilon,t} \leq 6 \exp\left(\frac{t}{2}(\|z^*\|_1 + 2\varepsilon)[\kappa_1(\|z^*\|_1 + 2\varepsilon) + \kappa_2] - \frac{\varepsilon\sqrt{V}}{6} e^{-t[\kappa_1(\|z^*\|_1 + 2\varepsilon) - \kappa_2]} \right),$$

where $\exp(h)$ is defined as e^h for all real numbers h.

A.2. Proof of Theorem 5.2. First of all, we define some quantities that are useful to give specific bounds on our approximation error. Define

$$\widetilde{\Lambda}_0^t = \max_{\widetilde{S} \in \widetilde{\mathcal{X}}} \sum_{\widetilde{S} + y \to \widetilde{S}' + y' \in \widetilde{\mathcal{R}}} \lambda_{\widetilde{S} + y \to \widetilde{S}' + y'}(\widetilde{S}, Z(t)),$$

$$\widetilde{L}_0^{\varepsilon,t} = \sup_{\substack{(z,z') \in \Omega_2^{\varepsilon,t} \\ z \neq z'}} \max_{\widetilde{S} \in \widetilde{\mathcal{X}}} \sum_{\widetilde{S} + y \to \widetilde{S}' + y' \in \widetilde{\mathcal{R}}} \frac{|\lambda_{\widetilde{S} + y \to \widetilde{S}' + y'}(\widetilde{S}, z) - \lambda_{\widetilde{S} + y \to \widetilde{S}' + y'}(\widetilde{S}, z')|}{\|z - z'\|_\infty}$$

$$\widetilde{\delta}_0^{\varepsilon,t} = \sup_{z \in \Omega_1^{\varepsilon,t}} \max_{\widetilde{S} \in \widetilde{\mathcal{X}}} \sum_{\widetilde{S} + y \to \widetilde{S}' + y' \in \widetilde{\mathcal{R}}} |\lambda^V_{\widetilde{S} + y \to \widetilde{S}' + y'}(\widetilde{S}, \lfloor Vz \rfloor) - \lambda_{\widetilde{S} + y \to \widetilde{S}' + y'}(\widetilde{S}, z)|$$

$$\widetilde{\Lambda}_1^t = \int_0^t \widetilde{\Lambda}_0^u du, \quad \widetilde{L}_1^{\varepsilon,t} = \int_0^t L_0^{\varepsilon,u} du, \quad \widetilde{\delta}_1^{\varepsilon,t} = \int_0^t \delta_0^{V,\varepsilon,u} du.$$

Note that $\widetilde{\Lambda}_0^t$ is finite for any $t \in [0,T]$, due to the fact that Z is defined over the whole interval $[0,T]$. Moreover the functions $\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}$ are locally Lipschitz on $\mathbb{R}^d_{>0}$ by Lemma 5.1, hence $\widetilde{L}_0^{\varepsilon,t}$ is finite for all $t \in [0,T]$. Finally, $\widetilde{\delta}_0^{V,\varepsilon,t}$ is finite for all $t \in [0,T]$ by Lemma 5.1. Note that, for fixed V and ε, the quantities $\widetilde{L}_0^{\varepsilon,t}$ and $\widetilde{\delta}_0^{\varepsilon,t}$ are non-decreasing functions of t. As a consequence, for all $t \in [0,T], \varepsilon \in \mathbb{R}_{>0}$, and $V \in \mathbb{Z}_{\geq 1}$ we have

$$\widetilde{\Lambda}_1^t \leq t\widetilde{\Lambda}_0^t, \quad \widetilde{L}_1^{\varepsilon,t} \leq t\widetilde{L}_0^{\varepsilon,t}, \quad \text{and} \quad \widetilde{\delta}_1^{\varepsilon,t} \leq t\widetilde{\delta}_0^{\varepsilon,t}. \tag{A.4}$$

Before proving Theorem 5.2 we show the following stronger result.

THEOREM A.4. *Assume that Assumption 3.1 holds. Furthermore, assume that the random variables $X^V(0)/V$ converge in probability to a constant z^* as V goes to infinity. Assume that the solution Z to (2.2) with $Z(0) = z^*$ exists over the interval $[0,T]$ and that*

$$m = \min_{\substack{S\in\mathcal{X}\\ u\in[0,T]}} Z_S(u) > 0.$$

Finally, assume that $Y^V(0) = Y(0)$ for all positive integers V. Then,

$$P\left(Y^V(t) \neq Y(t)\right) = E\left[\|Y^V(t) - Y(t)\|_\infty\right]. \tag{A.5}$$

Moreover, for any $0 < \varepsilon < m$

$$\sup_{t\in[0,T]} E\left[\|Y^V(t) - Y(t)\|_\infty\right] \leq p^{V,\varepsilon,T} + (\widetilde{\delta}_1^{V,\varepsilon,T} + \varepsilon\widetilde{L}_1^{\varepsilon,})e^{2\widetilde{\Lambda}_1^T}.$$

Proof. First, note that

$$\|Y^V(t) - Y(t)\|_\infty = \begin{cases} 1 & \text{if } Y^V(t) \neq Y(t) \\ 0 & \text{if } Y^V(t) = Y(t) \end{cases}, \tag{A.6}$$

hence (A.5) holds. Consider the process

$$\hat{Y}^V(t) = Y(0) + \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} (\widetilde{S}'-\widetilde{S}) N_{\widetilde{S}+y\to\widetilde{S}'+y'}\left(\int_0^t \lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\hat{Y}^V(u), X^{V,\varepsilon}(u))du\right). \tag{A.7}$$

By equations (5.2) and (A.7), using the triangular inequality, we obtain

$$\begin{aligned} &E\left[\|\hat{Y}^V(t) - Y(t)\|_\infty\right] \\ &\leq E\left[\int_0^t \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left|\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\hat{Y}^V(u), X^{V,\varepsilon}(u)) - \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(Y(u), Z(u))\right| du\right] \\ &\leq \Upsilon_1 + \Upsilon_2 + \Upsilon_3 \end{aligned}$$

where

$$\Upsilon_1 = E\left[\int_0^t \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left|\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\hat{Y}^V(u), X^{V,\varepsilon}(u)) - \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}\left(\hat{Y}^V(u), \frac{X^{V,\varepsilon}(u)}{V}\right)\right| du\right]$$

$$\Upsilon_2 = E\left[\int_0^t \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left|\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}\left(\hat{Y}^V(u), \frac{X^{V,\varepsilon}(u)}{V}\right) - \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(\hat{Y}^V(u), Z(u))\right| du\right]$$

$$\Upsilon_3 = E\left[\int_0^t \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left|\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(\hat{Y}^V(u), Z(u)) - \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(Y(u), Z(u))\right| du\right]$$

Since for every $\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}$ we have

$$\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(w,x) = \mathbb{1}_{\{\widetilde{S}\}}(w)\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S},x) \quad \text{for all } x\in\mathbb{Z}^d_{\geq 0}, w\in\widetilde{\mathcal{X}}$$
$$\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(w,z) = \mathbb{1}_{\{\widetilde{S}\}}(w)\lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S},z) \quad \text{for all } z\in\mathbb{R}^d_{\geq 0}, w\in\widetilde{\mathcal{X}},$$

we can write $\Upsilon_1 \leq \widetilde{\delta}_1^{V,\varepsilon,t}$. Similarly, $\Upsilon_2 \leq \varepsilon\widetilde{L}_1^{\varepsilon,t}$. Finally,

$$\begin{aligned}\Upsilon_3 &= E\left[\int_0^t \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left|\mathbb{1}_{\{\widetilde{S}\}}(\hat{Y}^V(u)) - \mathbb{1}_{\{\widetilde{S}\}}(Y(u))\right| \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S}, Z(u)) du\right]\\ &\leq E\left[\int_0^t \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}} \left|\mathbb{1}_{\{\widetilde{S}\}}(\hat{Y}^V(u)) - \mathbb{1}_{\{\widetilde{S}\}}(Y(u))\right| \widetilde{\Lambda}^u_0 du\right]\\ &= \int_0^t 2P\left(Y^V(u)\neq Y(u)\right)\widetilde{\Lambda}^u_0 du = 2\int_0^t E\Big[\|\hat{Y}^V(u) - Y(u)\|_\infty\Big]\widetilde{\Lambda}^u_0 du,\end{aligned}$$

where in the last equality we used (A.5). In conclusion,

$$E\Big[\|\hat{Y}^V(t) - Y(t)\|_\infty\Big] \leq (\widetilde{\delta}_1^{V,\varepsilon,t} + \varepsilon\widetilde{L}_1^{\varepsilon,t}) + 2\int_0^t E\Big[\|\hat{Y}^V(u) - Y(u)\|_\infty\Big]\widetilde{\Lambda}^u_0 du.$$

By the Gronwall inequality we then have

$$E\Big[\|\hat{Y}^V(t) - Y(t)\|_\infty\Big] \leq (\widetilde{\delta}_1^{V,\varepsilon,t} + \varepsilon\widetilde{L}_1^{\varepsilon,t})e^{2\widetilde{\Lambda}^t_1}.$$

The result follows by taking the sup over $t\in[0,T]$ on both sides (the quantity on the right-hand side of the inequality is non-decreasing in t) and by noting that $\mathbb{1}_{\mathcal{A}_{V,\varepsilon,T}}\hat{Y}^V(t) = \mathbb{1}_{\mathcal{A}_{V,\varepsilon,T}}Y^V(t)$ for all $t\in[0,T]$. Hence,

$$\begin{aligned}\|Y^V(t) - Y(t)\|_\infty &= \|Y^V(t) - Y(t)\|_\infty \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,T}} + \|\hat{Y}^V(t) - Y(t)\|_\infty \mathbb{1}_{\mathcal{A}_{V,\varepsilon,T}}\\ &\leq \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,T}} + \|\hat{Y}^V(t) - Y(t)\|_\infty \mathbb{1}_{\mathcal{A}_{V,\varepsilon,T}}\\ &\leq \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,T}} + \|\hat{Y}^V(t) - Y(t)\|_\infty.\end{aligned}$$

□

We are now ready to prove Theorem 5.2

Proof of Theorem 5.2. It follows from Theorem A.4 that $P\left(Y^V(t) \neq Y(t)\right) = E\left[\|Y^V(t) - Y(t)\|_\infty\right]$. Moreover, for any $\varepsilon > 0$ we have $\lim_{V\to\infty} p^{V,\varepsilon,T} = 0$ by Theorem 3.2, and $\lim_{V\to\infty} \widetilde{\delta}_1^{V,\varepsilon,T} = 0$ by Lemma 5.1 and (A.4). Hence,

$$\lim_{V\to\infty} \sup_{t\in[0,T]} E\left[\|Y^V(t) - Y(t)\|_\infty\right] \leq \varepsilon \widetilde{L}_1^{\varepsilon,T} e^{2\widetilde{\Lambda}_1^T},$$

which concludes the proof by the arbitrariness of $\varepsilon > 0$ and by the fact that $\widetilde{L}_0^{\varepsilon,T}$ (hence $\widetilde{L}_1^{\varepsilon,T}$) is non-decreasing in ε. ∎

A.3. Proof of Theorem 5.8. Similarly to what was done in the previous section, we define the following quantities to give an upper bound for our approximation error. Define

$$\hat{R} = \max_{y\to y'\in\mathcal{R}} \|\pi(y'-y)\|_\infty, \quad \hat{r} = \max_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left\| \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} - \frac{\tau(\widetilde{S})}{\alpha(\tau(\widetilde{S}))} \right\|_\infty,$$

$$\hat{\Lambda}_0^t = \hat{r} \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S}, Z(t)), \quad \hat{\Lambda}_1^t = \int_0^t \hat{\Lambda}_0^u du,$$

$$\hat{\Lambda}_2^t = \max_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \int_0^t \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S}, Z(u))du,$$

$$\hat{\Lambda}_3^{V,\varepsilon,t} = \int_0^t \sup_{z\in\Omega_1^{\varepsilon,u}} \sum_{y\to y'\in\mathcal{R}} \frac{\lambda_{y\to y'}^V(\lfloor Vz\rfloor)}{V} du,$$

$$\omega^{\varepsilon,t} = \hat{r} \sup_{\substack{(z,z')\in\Omega_2^{\varepsilon,t} \\ \|z-z'\|_\infty\leq\varepsilon}} \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \left| \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}\left(\widetilde{S}, z\right) - \lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S}, z') \right|,$$

$$\zeta^{\varepsilon,t} = \int_0^t (\|Z(u))\|_\infty + \varepsilon) du.$$

Note that $\hat{\Lambda}_0^t$, $\hat{\Lambda}_2^t$, and $\zeta^{\varepsilon,t}$ are finite for any $t \in [0,T]$, because Z is defined over the whole interval $[0,T]$ and the functions $\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}$ are continuous on $\mathbb{R}^d_{>0}$ by Lemma 5.1. Lemma 5.1 also implies that $\omega^{\varepsilon,t}$ is finite for all $t \in [0,T]$ and $\varepsilon \in \mathbb{R}_{>0}$. Finally, $\hat{\Lambda}_3^{V,\varepsilon,t}$ is finite by Assumption 3.1. Note that, for fixed V and ε, the quantities $\hat{\Lambda}_3^{V,\varepsilon,t}$, $\omega^{\varepsilon,t}$, and $\zeta^{\varepsilon,t}$ are non-decreasing functions of t.

We now state and prove the following result, which immediately implies Theorem 5.8. Note that $\delta_1^{V,\varepsilon,t}$ is as defined in Section A.1.

THEOREM A.5. *Consider a family of stochastic reaction systems with trackable species (Y^V, X^V), and assume that Assumptions 3.1 and 5.6 are satisfied. Let $z^* \in \mathbb{R}^d_{>0}$ and $\widetilde{X}^V(0) = \lfloor Vz^*\rfloor$. Define the process $\widetilde{X}^V$ by*

$$\widetilde{X}^V(t) = \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)} \frac{\tau(Y^{\widetilde{S},i}(t))}{\alpha(\tau(Y^{\widetilde{S},i}(t)))},$$

where the processes $(Y^{\widetilde{S},i})_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}, i\in\mathbb{Z}_{\geq 1}}$ are independent and satisfy

$$Y^{\widetilde{S},i}(t) = \widetilde{S} + \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} (\widetilde{S}''-\widetilde{S}') N^{\widetilde{S},i}_{\widetilde{S}'+y\to\widetilde{S}''+y'} \left(\int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y(u)^{\widetilde{S},i}, Z(u))du \right),$$

for a family of independent, identically distributed unit-rate Poisson processes $\{N_r^{\widetilde{S},i}\}_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\},i\in\mathbb{Z}_{\geq 1},r\in\widetilde{\mathcal{R}}}$. *For arbitrary* $\nu_1,\nu_2,\nu_3\in\mathbb{R}_{>0}$ *define*

$$\nu = e^{\hat{\Lambda}_1^T}\left(\hat{R}\nu_1 + \hat{r}\nu_2 + \nu_3 + \hat{R}\delta_1^{V,\varepsilon,T} + \omega^{\varepsilon,T}\zeta^{\varepsilon,T}\right)$$

Then,

$$P\left(\sup_{0\leq t\leq T}\left\|\frac{\pi(X^V(t))}{V} - \frac{\widetilde{X}^V(t)}{V}\right\|_\infty > \nu\right) \leq 6\exp\left(\frac{e\hat{\Lambda}_3^{V,\varepsilon,t}}{2} - \frac{\nu_1\sqrt{V}}{3}\right)$$
$$+ 6\exp\left(\frac{ec\hat{\Lambda}_2^t}{2} - \frac{\nu_2\sqrt{V}}{3}\right) + P\left(\left\|\frac{\pi(X^V(0))}{V} - \frac{\widetilde{X}^V(0)}{V}\right\|_\infty > \nu_3\right) + p^{V,\varepsilon,T},$$

where $c = \sum_{S\in\mathcal{X}}\alpha(S)z_S^*$.

Proof. By the superposition property of Poisson processes, for all $V\in\mathbb{Z}_{\geq 1}$ there exist two unit-rate Poisson processes U_1^V and U_2^V such that for all $t\in\mathbb{R}_{\geq 0}$

$$U_1^V\left(\sum_{y\to y'\in\mathcal{R}}\int_0^t \lambda_{y\to y'}^V(X^{V,\varepsilon}(u))du\right) = \sum_{y\to y'\in\mathcal{R}} N_{y\to y'}\left(\int_0^t \lambda_{y\to y'}^V(X^{V,\varepsilon}(u))du\right)$$

and

$$U_2^V\left(\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(u),Z(u))du\right)$$
$$= \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} N^{\widetilde{S},i}_{\widetilde{S}'+y\to\widetilde{S}''+y'}\left(\int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(u),Z(u))du\right)$$

Note that

$$\widetilde{X}^V(t) = \widetilde{X}^V(0) + \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\left(\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))} - \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right)\times$$
$$\times N^{\widetilde{S},i}_{\widetilde{S}'+y\to\widetilde{S}''+y'}\left(\int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y(u)^{\widetilde{S},i},Z(u))du\right).$$

Hence, by triangular inequality,

$$\sup_{0\leq u\leq t}\left\|\frac{\pi(\hat{X}^{V,\varepsilon}(u))}{V} - \frac{\widetilde{X}^V(u)}{V}\right\|_\infty \leq \left\|\frac{\pi(X^V(0))}{V} - \frac{\widetilde{X}^V(0)}{V}\right\|_\infty + \sum_{i=1}^5\Upsilon_i$$

where

$$\Upsilon_1 = \sup_{0\le u\le t} \sum_{y\to y'\in\mathcal{R}} \|\pi(y'-y)\|_\infty \frac{1}{V}\left|\overline{N}_{y\to y'}\left(\int_0^u \lambda^V_{y\to y'}(X^{V,\varepsilon}(w))dw\right)\right|$$

$$\le \frac{\hat{R}}{V}\sup_{0\le u\le t}\left|\overline{U}^V_1\left(\sum_{y\to y'\in\mathcal{R}}\int_0^u \lambda^V_{y\to y'}(X^{V,\varepsilon}(w))dw\right)\right|$$

$$\Upsilon_2 = \sup_{0\le u\le t}\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\left\|\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))}-\frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right\|_\infty \times$$

$$\times\frac{1}{V}\left|\overline{N}^{\widetilde{S},i}_{\widetilde{S}'+y\to\widetilde{S}''+y'}\left(\int_0^u \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(w),Z(w))dw\right)\right|$$

$$\le \frac{\hat{r}}{V}\sup_{0\le u\le t}\left|\overline{U}^V_2\left(\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\int_0^u \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(w),Z(w))dw\right)\right|$$

$$\Upsilon_3 = \sup_{0\le u\le t}\sum_{y\to y'\in\mathcal{R}}\|\pi(y'-y)\|_\infty\int_0^u\left|\frac{\lambda^V_{y\to y'}(X^{V,\varepsilon}(w))}{V}-\lambda_{y\to y'}\left(\frac{X^{V,\varepsilon}(w)}{V}\right)\right|dw$$

$$\le \hat{R}\delta^{V,\varepsilon,t}_1$$

$$\Upsilon_4 = \sup_{0\le u\le t}\left\|\sum_{y\to y'\in\mathcal{R}}\pi(y'-y)\int_0^u\lambda_{y\to y'}\left(\frac{X^{V,\varepsilon}(w)}{V}\right)dw\right.$$

$$\left.-\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\left(\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))}-\frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right)\int_0^u\frac{X^{V,\varepsilon}_{\tau(\widetilde{S}')}(w)}{V}\lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(\widetilde{S}',Z(w))dw\right\|_\infty$$

$$\Upsilon_5 = \sup_{0\le u\le t}\left\|\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\left(\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))}-\frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right)\int_0^u\frac{X^{V,\varepsilon}_{\tau(\widetilde{S}')}(w)}{V}\lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(\widetilde{S}',Z(w))dw\right.$$

$$\left.-\frac{1}{V}\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\left(\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))}-\frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right)\int_0^u\lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(w),Z(w))dw\right\|_\infty$$

We first focus on rewriting Υ_4 and Υ_5. To this aim, first note that by identifying species with canonical vectors of $\mathbb{R}^d$ as previously done in the paper, we have that for all $y\in\mathcal{C}$

$$\pi(y)=\sum_{S\in\overline{\mathcal{X}}}y_S S=\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\frac{y_{\tau(\widetilde{S})}\tau(\widetilde{S})}{\alpha(\tau(\widetilde{S}))}.$$

Hence, for all $y \to y' \in \mathcal{R}$

$$\begin{aligned}\pi(y'-y) &= \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \frac{y_{\tau(\widetilde{S}')}\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} - \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \frac{y_{\tau(\widetilde{S})}\tau(\widetilde{S})}{\alpha(\tau(\widetilde{S}))} \\ &= \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} p_{y\to y'}(\widetilde{S},\widetilde{S}') - \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \frac{y_{\tau(\widetilde{S})}}{\alpha(\tau(\widetilde{S}))}\tau(\widetilde{S}),\end{aligned}$$

where we used Assumption 5.6 in the last equality. By recalling that $\tau(\Delta) = 0$ and $\sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}} p_{y\to y'}(\widetilde{S},\widetilde{S}')$ for all $y \to y' \in \mathcal{R}$ and $\widetilde{S} \in \widetilde{\mathcal{X}}$, we further obtain

$$\begin{aligned}\pi(y'-y) &= \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}} \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} y_{\tau(\widetilde{S})} p_{y\to y'}(\widetilde{S},\widetilde{S}') \\ &\quad - \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \frac{y_{\tau(\widetilde{S})}}{\alpha(\tau(\widetilde{S}))}\tau(\widetilde{S}) \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}} p_{y\to y'}(\widetilde{S},\widetilde{S}') \\ &= \sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{\widetilde{S}'\in\widetilde{\mathcal{X}}} \left(\frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} - \frac{\tau(\widetilde{S})}{\alpha(\tau(\widetilde{S}))}\right) y_{\tau(\widetilde{S})} p_{y\to y'}(\widetilde{S},\widetilde{S}').\end{aligned}$$

It follows that

$$\begin{aligned}&\sum_{y\to y'\in\mathcal{R}} \pi(y'-y) \int_0^u \lambda_{y\to y'}\left(\frac{X^{V,\varepsilon}(w)}{V}\right) dw \\ &= \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} \left(\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))} - \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right) \int_0^u y_{\tau(\widetilde{S}')} p_{y\to y'}(\widetilde{S}',\widetilde{S}'') \lambda_{y\to y'}\left(\frac{X^{V,\varepsilon}(w)}{V}\right) dw \\ &= \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} \left(\frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))} - \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))}\right) \int_0^u \frac{X^{V,\varepsilon}_{\tau(\widetilde{S})}(w)}{V} \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}\left(\widetilde{S}', \frac{X^{V,\varepsilon}(w)}{V}\right) dw,\end{aligned}$$

which in turn implies

$$\begin{aligned}\Upsilon_4 &\le \sup_{0\le u\le t} \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} \left\| \frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))} - \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} \right\|_\infty \times \\ &\quad \times \int_0^u \frac{X^{V,\varepsilon}_{\tau(\widetilde{S})}(w)}{V} \left| \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}\left(\widetilde{S}', \frac{X^{V,\varepsilon}(w)}{V}\right) - \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(\widetilde{S}', Z(w)) \right| dw \\ &\le \omega^{\varepsilon,t}\zeta^{\varepsilon,t}.\end{aligned}$$

By summing over the values of the single-molecule trajectories, we also have

$$\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)} \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(w), Z(w)) = \widetilde{X}^V_{\tau(\widetilde{S})}(w) \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(\widetilde{S}, Z(w)),$$

which implies

$$\Upsilon_5 \leq \sup_{0\leq u\leq t} \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} \left\| \frac{\tau(\widetilde{S}'')}{\alpha(\tau(\widetilde{S}''))} - \frac{\tau(\widetilde{S}')}{\alpha(\tau(\widetilde{S}'))} \right\|_\infty \int_0^u \left| \frac{X^{V,\varepsilon}_{\tau(\widetilde{S})}(w)}{V} - \frac{\widetilde{X}^V_{\tau(\widetilde{S})}(w)}{V} \right| \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(\widetilde{S}', Z(w))dw$$

$$\leq \int_0^t \left\| \frac{X^{V,\varepsilon}(u)}{V} - \frac{\widetilde{X}^V(u)}{V} \right\|_\infty \hat{\Lambda}_0^u du$$

$$= \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,t}} \int_0^t \left\| \frac{X^{V,\varepsilon}(u)}{V} - \frac{\widetilde{X}^V(u)}{V} \right\|_\infty \hat{\Lambda}_0^u du + \mathbb{1}_{\mathcal{A}_{V,\varepsilon,t}} \int_0^t \left\| \frac{\hat{X}^{V,\varepsilon}(u)}{V} - \frac{\widetilde{X}^V(u)}{V} \right\|_\infty \hat{\Lambda}_0^u du.$$

$$\leq \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,t}} M^{V,\varepsilon,t} + \int_0^t \left\| \frac{\hat{X}^{V,\varepsilon}(u)}{V} - \frac{\widetilde{X}^V(u)}{V} \right\|_\infty \hat{\Lambda}_0^u du,$$

where

$$M^{V,\varepsilon,t} = \int_0^t \left(\|Z(u)\|_\infty + \varepsilon + \sum_{S\in\mathcal{X}} \alpha(S) \frac{\widetilde{X}^V_S(0)}{V} \right) \hat{\Lambda}_0^u du$$

is an almost surely finite random variable, non-decreasing in t. Hence, putting everything together and applying the Gronwall inequality we have that almost surely

$$\sup_{0\leq t\leq T} \left\| \frac{\pi(\hat{X}^{V,\varepsilon}(t))}{V} - \frac{\widetilde{X}^V(t)}{V} \right\|_\infty \leq e^{\hat{\Lambda}_1^T} \frac{\hat{R}}{V} \sup_{0\leq t\leq T} \left| \overline{U}_1^V \left(\sum_{y\to y'\in\mathcal{R}} \int_0^t \lambda^V_{y\to y'}(X^{V,\varepsilon}(u))du \right) \right|$$

$$+ e^{\hat{\Lambda}_1^T} \frac{\hat{r}}{V} \sup_{0\leq t\leq T} \left| \overline{U}_2^V \left(\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} \sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)} \int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(u), Z(u))du \right) \right|$$

$$+ e^{\hat{\Lambda}_1^T} \left(\left\| \frac{\pi(X^V(0))}{V} - \frac{\widetilde{X}^V(0)}{V} \right\|_\infty + \hat{R}\delta_1^{V,\varepsilon,T} + \omega^{\varepsilon,T}\zeta^{\varepsilon,T} + \mathbb{1}_{\mathcal{A}^c_{V,\varepsilon,T}} M^{V,\varepsilon,T} \right).$$

Now note that if $A_1, A_2, \ldots, A_j$ are random variables and $a_1, a_2, \ldots, a_j$ are positive real numbers, then

$$P\left(\sum_{i=1}^j A_i > \sum_{i=1}^j a_i \right) \leq P\left(\bigcup_{i=1}^j (A_i > a_i) \right) \leq \sum_{i=1}^j P(A_i > a_i).$$

Hence, if ν is as in the statement of the theorem and $\nu < \varepsilon$,

$$P\left(\sup_{0\leq t\leq T} \left\| \frac{\pi(X^V(t))}{V} - \frac{\widetilde{X}^V(t)}{V} \right\|_\infty > \nu \right) = P\left(\sup_{0\leq t\leq T} \left\| \frac{\pi(\hat{X}^{V,\varepsilon}(t))}{V} - \frac{\widetilde{X}^V(t)}{V} \right\|_\infty > \nu \right)$$

$$\leq P\left(\frac{1}{V} \sup_{0\leq t\leq T} \left| \overline{U}_1^V \left(\sum_{y\to y'\in\mathcal{R}} \int_0^t \lambda^V_{y\to y'}(X^{V,\varepsilon}(u))du \right) \right| > \nu_1 \right)$$

$$+ P\left(\frac{1}{V} \sup_{0\leq t\leq T} \left| \overline{U}_2^V \left(\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}} \sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}} \sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)} \int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(u), Z(u))du \right) \right| > \nu_2 \right)$$

$$+ p^{V,\varepsilon,T}.$$

Since for all $t \in [0,T]$

$$\int_0^t \lambda^V_{y\to y'}(X^{V,\varepsilon}(u))du \leq V\hat{\Lambda}_3^{V,\varepsilon,t}$$

and

$$\sum_{\widetilde{S}\in\widetilde{\mathcal{X}}\setminus\{\Delta\}}\sum_{\widetilde{S}'+y\to\widetilde{S}''+y'\in\widetilde{\mathcal{R}}}\sum_{i=1}^{\widetilde{X}^V_{\tau(\widetilde{S})}(0)}\int_0^t \lambda_{\widetilde{S}'+y\to\widetilde{S}''+y'}(Y^{\widetilde{S},i}(u),Z(u))du \leq Vc\hat{\Lambda}_2^t,$$

the proof is concluded by Lemma A.1. ∎

Proof of Theorem 5.8. Note that by Lemma 5.7 and by the fact that $\alpha(S) \geq 1$ for all $S \in \mathcal{X}$ in (5.6),

$$\begin{aligned}\left\|\frac{\pi(X^V(h))}{V} - \frac{\widetilde{X}^V(h)}{V}\right\|_1 &\leq \left\|\frac{\pi(X^V(h))}{V}\right\|_1 + \left\|\frac{\widetilde{X}^V(h)}{V}\right\|_1 \\ &\leq \frac{1}{V}\left(\sum_{S\in\overline{\mathcal{X}}}\alpha(S)\left(X^V_S(0)+\widetilde{X}^V_S(0)\right)\right).\end{aligned}$$

Under the assumption that both $X^V(0)$ and $\widetilde{X}^V(0)$ have finite expectation and converge in probability to z^*, and by the equivalence of norms in finite dimension, we conclude there exists $M \in \mathbb{R}_{>0}$ such that

$$\sup_{V\in\mathbb{Z}_{\geq 1}} E\left[\left\|\frac{\pi(X^V(h))}{V} - \frac{\widetilde{X}^V(h)}{V}\right\|_\infty\right] \leq M.$$

Hence, if ν is as in Theorem A.5, we have that

$$\begin{aligned}E\left[\sup_{0\leq t\leq T}\left\|\frac{\pi(X^V(h))}{V} - \frac{\widetilde{X}^V(h)}{V}\right\|_\infty\right] &\leq \nu + 6Me^{\frac{\hat{\Lambda}_3^{V,\varepsilon,t}}{2}-\frac{\nu_1\sqrt{V}}{3}} \\ &+ 6Me^{\frac{c\hat{\Lambda}_2^t}{2}-\frac{\nu_2\sqrt{V}}{3}} + MP\left(\left\|\frac{\pi(X^V(0))}{V} - \frac{\widetilde{X}^V(0)}{V}\right\|_\infty > \nu_3\right) + Mp^{V,\varepsilon,T}.\end{aligned}$$

The proof is concluded if we can show that for all $T \in \mathbb{R}_{>0}$ and any arbitrary $\eta > 0$, we can fix $\nu_1,\nu_2,\nu_3 \in \mathbb{R}_{>0}$ and $\varepsilon \in (0,m)$ such that $\nu < \eta$ for large enough values of V. Indeed, for any fixed $\varepsilon \in (0,m), T \in \mathbb{R}_{>0}$ the other terms on the right-hand side of the above inequality tend to zero as V goes to infinity. To show that ν can be made smaller than η, simply note that ν_1,ν_2,ν_3 can be chosen as small as desired among the positive real numbers, $\delta_1^{V,\varepsilon,T}$ tends to zero as V goes to infinity for all fixed $\varepsilon \in (0,m)$ by Assumption 3.1, and $\omega^{\varepsilon,T}$ tends to zero as ε tends to zero because the functions $\lambda_{\widetilde{S}+y\to\widetilde{S}'+y'}$ are locally Lipschitz on $\widetilde{\mathcal{X}}\times\mathbb{R}^d_{>0}$ by Lemma 5.1. ∎

A.4. Proof of Theorem 5.5. Note that under the assumptions of Theorem 5.5, for all $t \in [0,T]$ $Y^V(t)$ converges in probability to $Y(t)$ by Theorem 5.2. Hence, in order to prove Theorem 5.5, we need to show relative compactness of $\{Y^V\}$ as a sequence of processes with sample paths in $D_{\widetilde{\mathcal{X}}}[0,T]$, and conclude by [10, Lemma A2.1]. To prove relative compactness of $\{Y^V\}$, first note that the state space $\widetilde{\mathcal{X}}$ is

compact. Hence, we only need to show that the jump times do not accumulate as V tends to infinity, nor tend to 0. Let t_i^V with $i \in \mathbb{Z}_{\geq 1}$ denote the time of the ith jump of Y^V, let $t_0^V = 0$, and let T^V be the time of the last jump of Y^V in $[0,T]$. Fix $\delta \in \mathbb{R}_{>0}$ and for all $j \in \mathbb{Z}$ with $-1 \leq j \leq T/\delta$ let $N_j^{V,\delta}$ be the number of jumps of Y^V in the interval $[j/\delta, \min\{j/\delta + 2\delta, T\}]$. The $N_j^{V,\delta}$ are introduced to control the time between jumps: whenever two jumps occur at times closer than δ, there necessarily exists an interval $[j/\delta, \min\{j/\delta + 2\delta, T\}]$ with $j \geq 0$ containing both of them. Also, whenever the time of a jump is smaller than 0, then $N_{-1}^{V,\delta} \geq 1$. Hence, for all $\varepsilon \in \mathbb{R}_{>0}$ with $\varepsilon > m$,

$$
\begin{aligned}
P\left(\min_{j=1,\dots,T^V}(t_j^V - t_{j-1}^V) \leq \delta\right) &\leq P\left(N_{-1}^{V,\delta} \geq 1 \text{ or } \max_{j=1,\dots,\lfloor T/\delta \rfloor} N_j^{V,\delta} \geq 2\right) \\
&\leq P\left(N_{-1}^{V,\delta} \geq 1\right) + \sum_{j=1}^{\lfloor T/\delta \rfloor} P(N_j^{V,\delta} \geq 2) \\
&\leq P\left(\sup_{0\leq t\leq T} \left\| \frac{X^V}{V}(t) - Z(t) \right\|_\infty > \varepsilon\right) + P(N^\varepsilon(\delta) \geq 1) + \frac{T}{\delta} P(N^\varepsilon(2\delta) \geq 2),
\end{aligned}
$$

where N^ε is a Poisson process with rate

$$
B_\varepsilon = \sup_{N\in\mathbb{Z}_{\geq 1}} \sup_{z\in\Omega_1^{\varepsilon,T}} \max_{\widetilde{S}\in\widetilde{\mathcal{X}}} \sum_{\widetilde{S}+y\to\widetilde{S}'+y'\in\widetilde{\mathcal{R}}} \lambda^V_{\widetilde{S}+y\to\widetilde{S}'+y'}(\widetilde{S}, \lfloor Vz \rfloor),
$$

which is finite by Lemma 5.1. Hence, by Theorem 3.2

$$
\limsup_{V\to\infty} P\left(\min_{j=1,\dots,T^V}(t_j^V - t_{j-1}^V) \leq \delta\right) \leq (1 - e^{-\delta B_\varepsilon}) + \frac{T}{\delta}(1 - e^{-2\delta B_\varepsilon} - 2\delta B_\varepsilon e^{-2\delta B_\varepsilon}),
$$

which tends to 0 as δ tends to 0. Therefore, $\{Y^V\}$ is relatively compact as a sequence of processes with sample paths in $D_{\widetilde{\mathcal{X}}}[0,T]$ by [11, Corollary 7.4, Chapter 3], which completes the proof.

REFERENCES

[1] A. Agazzi, L. Andreis, R. I. Patterson, and D. M. Renger, *Large deviations for markov jump processes with uniformly diminishing rates*, Stochastic Processes and their Applications, 152 (2022), pp. 533–559.

[2] A. Agazzi, A. Dembo, and J.-P. Eckmann, *Large deviations theory for markov jump models of chemical reaction networks*, The Annals of Applied Probability, 28 (2018), pp. 1821–1855.

[3] D. F. Anderson, D. Cappelletti, J. Kim, and T. D. Nguyen, *Tier structure of strongly endotactic reaction networks*, Stochastic Processes and their Applications, 130 (2020), pp. 7218–7259.

[4] D. F. Anderson, A. Ganguly, and T. G. Kurtz, *Error analysis of tau-leap simulation methods*, The Annals of Applied Probability, 21 (2011), pp. 2226 – 2262.

[5] D. F. Anderson, D. J. Higham, S. C. Leite, and R. J. Williams, *On constrained langevin equations and (bio) chemical reaction networks*, Multiscale Modeling & Simulation, 17 (2019), pp. 1–30.

[6] D. F. Anderson and T. G. Kurtz, *Stochastic analysis of biochemical systems*, Springer, 2015.

[7] C. D. Bastian and G. A. Rempala, *Throwing stones and collecting bones: Looking for poisson-like random measures*, Mathematical Methods in the Applied Sciences, 43 (2020), pp. 4658–4668.

[8] W. R. K. Bukhsh, C. D. Bastian, M. Wascher, C. Klaus, S. Y. Sahai, M. H. Weir, E. Kenah, E. Root, J. H. Tien, and G. A. Rempala, *Projecting COVID-19 Cases and Subsequent Hospital Burden in Ohio*, medRxiv, (2022).

[9] F. Di Lauro, W. R. KhudaBukhsh, I. Z. Kiss, E. Kenah, M. Jensen, and G. A. Rempała, *Dynamic survival analysis for non-markovian epidemic models*, Journal of the Royal Society Interface, 19 (2022), p. 20220124.
[10] P. Donnelly and T. G. Kurtz, *A countable representation of the Fleming-Viot measure-valued diffusion*, The Annals of Probability, 24 (1996), pp. 698 – 742.
[11] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*, John Wiley & Sons Inc, 1986.
[12] D. T. Gillespie, *Approximate accelerated stochastic simulation of chemically reacting systems*, The Journal of Chemical Physics, 115 (2001), pp. 1716–1733.
[13] H.-W. Kang, T. G. Kurtz, and L. Popovic, *Central limit theorems and diffusion approximations for multiscale markov chain models*, The Annals of Applied Probability, 24 (2014), pp. 721–759.
[14] W. KhudaBukhsh, S. K. Khalsa, E. Kenah, G. Rempala, and J. H. Tien, *Covid-19 dynamics in an ohio prison*, medRxiv, (2021).
[15] W. R. KhudaBukhsh, B. Choi, E. Kenah, and G. A. Rempała, *Survival dynamical systems: individual-level survival analysis from population-level epidemic models*, Interface focus, 10 (2020), p. 20190048.
[16] T. G. Kurtz, *The relationship between stochastic and deterministic models for chemical reactions*, The Journal of Chemical Physics, 57 (1972), pp. 2976–2978.
[17] T. G. Kurtz, *Limit theorems and diffusion approximations for density dependent markov chains*, in Stochastic Systems: Modeling, Identification and Optimization, I, Springer, 1976, pp. 67–78.
[18] P. Mozgunov, M. Beccuti, A. Horvath, T. Jaki, R. Sirovich, and E. Bibbona, *A review of the deterministic and diffusion approximations for stochastic chemical reaction networks*, Reaction Kinetics, Mechanisms and Catalysis, 123 (2018), pp. 289–312.
[19] J. R. Norris, *Markov chains*, Cambridge university press, 1998.
[20] A. Prodhomme, *Strong gaussian approximation of metastable density-dependent markov chains on large time scales*, arXiv preprint arXiv:2010.06861, (2020).
[21] I. Somekh, W. R. KhudaBukhsh, E. D. Root, L. K. Boker, G. Rempala, E. A. Simões, and E. Somekh, *Quantifying the population-level effect of the covid-19 mass vaccination campaign in israel: a modeling study*, in Open forum infectious diseases, vol. 9, Oxford University Press US, 2022, p. ofac087.
[22] E. Sontag, *Structure and stability of certain chemical networks and applications to the kinetic proofreading model of t-cell receptor signal transduction*, IEEE Transactions on Automatic Control, 46 (2001), pp. 1028 – 1047.
[23] M. Wascher, P. M. Schnell, W. R. Khudabukhsh, M. Quam, J. H. Tien, and G. A. Rempala, *Monitoring sars-cov-2 transmission and prevalence in populations under repeated testing*, medRxiv, (2021).