Ergodicity analysis and antithetic integral control of a class of stochastic reaction networks with delays

Corentin Briat and Mustafa Khammash*
D-BSSE, ETH-Zürich

Abstract

Delays are an important phenomenon arising in a wide variety of real world systems. They occur in biological models because of diffusion effects or as simplifying modeling elements. We propose here to consider delayed stochastic reaction networks. The difficulty here lies in the fact that the state-space of a delayed reaction network is infinite-dimensional, which makes their analysis more involved. We demonstrate here that a particular class of stochastic time-varying delays, namely those that follow a phase-type distribution, can be exactly implemented in terms of a chemical reaction network. Hence, any delay-free network can be augmented to incorporate those delays through the addition of delay-species and delay-reactions. Hence, for this class of stochastic delays, which can be used to approximate any delay distribution arbitrarily accurately, the state-space remains finite-dimensional and, therefore, standard tools developed for standard reaction network still apply. In particular, we demonstrate that for unimolecular mass-action reaction networks that the delayed stochastic reaction network is ergodic if and only if the non-delayed network is ergodic as well. Bimolecular reactions are more difficult to consider but an analogous result is also obtained. These results tell us that delays that are phase-type distributed, regardless of their distribution, are not harmful to the ergodicity property of reaction networks. We also prove that the presence of those delays adds convolution terms in the moment equation but does not change the value of the stationary means compared to the delay-free case. Finally, the control of a certain class of delayed stochastic reaction network using a delayed antithetic integral controller is considered. It is proven that this controller achieves its goal provided that the delay-free network satisfy the conditions of ergodicity and output-controllability.

Keywords. Stochastic reaction networks; delay systems; ergodicity analysis; antithetic integral control.

*email: corentin@briat.info, mustafa.khammash@bsse.ethz.ch; url: www.briat.info, https://www.bsse.ethz.ch/ctsb.

1 Introduction

Delays are important physical phenomena induced by memory, propagation or transport effects [1–6]. They naturally arise in population dynamics [7], ecology [8], epidemiology [9], biology [10–14] and engineering [2, 3, 5, 15]. It is commonly understood that delays have, in general, detrimental effects in engineering as they may lead to instabilities such as oscillations. While this destabilizing effect is undesirable in this setting, their role can be crucial in biology when one wants to design oscillators [10, 16, 17]. In the stochastic setting, delays have indeed been shown to be helpful for generating oscillations [10], but also to accelerate signaling [18] or to be responsible for an increase of intrinsic variability [19]. Delays can be easily incorporated in the dynamics of a deterministic reaction network by simply substituting delay-free terms by delayed ones, thereby turning ordinary differential equations into delay-differential equations, the analysis of which can be readily carried out using well-developed techniques such as Lyapunov-based ones or input-output methods; see e.g. [1–5]. When the dynamics of the reaction network is inherently stochastic and represented by a continuous-time jump Markov process [20], the introduction of constant deterministic delays in the dynamics is also possible. Those networks can be easily simulated using a simple adaptation of Gillespie's stochastic simulation algorithm [10], the next reaction method [21] or using delayed continuous-time Markov chains [22]. Alternatively, it has been shown in [23–25] that certain chains of stochastic unimolecular reactions could be substituted by a stochastic time-varying delay whose distribution can exactly computed from the reaction rates. This has led to drastic improvements of the stochastic simulation algorithm, notably in terms of simulation speed.

When stochastic reaction networks modeled as jump Markov processes are considered, it has been shown that the notion ergodicity is a natural notion of stability that can be established using algebraic, graph theoretical and optimization techniques [26–31]. Ergodicity is the stochastic analogue of having a unique globally attractive fixed point for deterministic dynamics. It can also be used to establish moment convergence as well as the property that the population behavior can be deduced from a single trajectory of the Markov process. Checking whether a stochastic reaction network is ergodic amounts at showing two properties: the irreducibility of the state-space (or a subset of it) and the fulfilment of a Lyapunov-Foster condition. While it is quite clear how these conditions could be checked for standard (i.e. undelayed) stochastic reaction networks, the case of delayed stochastic reaction networks is more complicated. Indeed, since the state-space of a general delayed reaction network is infinite-dimensional, checking the irreducibility of a function-space is way more involved. The Foster-Lyapunov condition which is based on the use of a norm-like function is also more complicated to consider and tools from time-delay systems theory, such as Lyapunov-Krasovskii functionals, may need to be used.

The objective of this paper is to develop a framework for the modeling, the analysis and the control of reaction networks with stochastically time-varying delays. However, our goal is to avoid to redevelop a complete theory for the ergodicity analysis of Markov jump processes with infinite-dimensional state-space. Interestingly, this can be done by assuming that the delays follow a phase-type distribution, a class of distribution arising, for instance, in

queueing networks [32], risk theory [33], health-care [34] and evolution [35]. As those distributions are dense (in the sense of weak convergence) in the set of all probability distributions on $(0, \infty)$ [36], they can be used to approximate any delay-distribution with arbitrary precision. Several algorithms are available for approximating a given distribution or for fitting empirical ones; see e.g. [37]. The reason behind the use of such distributions is that they can be exactly modeled as an irreducible unimolecular reaction network themselves which means that delays can be included in the network by suitably adding extra species and extra reactions, thereby keeping the dimension of the state finite. As a consequence, existing tools can be applied to the augmented network to yield results on delayed reaction networks when the delays are phase-type distributed.

We, therefore, propose to use the tools developed in [26, 38] in order to establish several results for delayed reaction networks. Using the ergodicity results developed in [26], we prove here that a delayed unimolecular networks is ergodic if and only if its delay-free counterpart is ergodic as well. This result is interesting for two reasons. The first one is that phase-type distributed delays are harmless in the context of unimolecular networks. The second one is that the network will remain ergodic for any phase-type distributed delays, regardless of its complexity. This includes complex distributions approximating arbitrarily closely heavy-tailed distributions or even Dirac distributions. It is shown that similar results hold true in the context of bimolecular reaction networks.

The study of the moments equation of the delayed reaction network allows to show that the presence of the delays introduce convolution terms in the dynamics of the species of the reaction networks. The convolution kernels coincide with the delay-distributions. Interestingly, we also show that the stationary value for the mean copy number of the species is the same as for the delay-free network.

Finally, we address the problem of controlling delayed reaction networks using the antithetic integral controller proposed in [38]. We notably generalize the controller to include delayed reactions, in particular, delays are included in the actuation and the measurement reaction. We show that the delayed reaction network satisfy the ergodicity and output-controllability conditions if and only if the delay-free reaction satisfies the very same conditions. This result parallels those obtained for the ergodicity analysis. In this regard, if the delay-free network verifies the ergodicity and output-controllability conditions then the delayed network will also verify them, regardless of the complexity of the delay distributions.

Outline. Some preliminaries on reaction networks are first given in Section 2. Phase-type distributed delays are introduced in Section 3 and fully characterized in terms of algebraic conditions. A constructive procedure for building the associated reaction network is also provided. Section 4 introduces delayed reaction with phase-type distributed delays. Ergodicity conditions for those networks are provided in Section 5. The associated moments equation is briefly studied in Section 6. Finally, conditions ensuring their control using an antithetic integral controller are obtained in Section 7. Concluding discussions are provided in Section 8.

Notations. The cones of positive and nonnegative d-dimensional vectors are denoted by $\mathbb{R}^d_{>0}$ and $\mathbb{R}^d_{\geq 0}$, respectively, whereas the set of nonnegative integers is denoted by $\mathbb{Z}_{\geq 0}$. The vector $\mathbb{1}$ is the vector of ones. The operators $\mathrm{diag}_i(x_i) = \mathrm{diag}(x_1, \ldots, x_n)$, $\mathrm{col}_i(x_i) = \mathrm{col}(x_1, \ldots, x_n)$ and $\mathrm{row}_i(x_i) = \mathrm{row}(x_1, \ldots, x_n)$ denote the matrices consisting of placing the elements diagonally, vertically and horizontally, respectively. A square matrix $M \in \mathbb{R}^{d\times d}$ is said to be Hurwitz stable (or simply Hurwitz or stable) if all its eigenvalues have negative real part. The matrix M is said to be Metzler if all its off-diagonal elements are nonnegative.

2 Preliminaries on stochastic reaction networks

2.1 Stochastic reaction networks without delays

A reaction network $(\boldsymbol{X}, \mathcal{R})$ is a set of d molecular species $\boldsymbol{X} = (\boldsymbol{X_1}, \ldots, \boldsymbol{X_d})$ interacting via K reaction channels $\mathcal{R} := \{\mathcal{R}_1, \ldots, \mathcal{R}_K\}$. For each of reaction, we denote the stoichiometric vector of the k-th reaction by $\zeta_k \in \mathbb{Z}^d$ and the propensity of the k-th reaction by $\lambda_k(\cdot)$ where $\lambda_k : \mathbb{Z}^d_{\geq 0} \to \mathbb{R}_{\geq 0}$ with the additional condition that if $x + \zeta_k \notin \mathbb{Z}^d_{\geq 0}$ then $\lambda_k(x) = 0$. Under the well-mixed assumption, the process $(X(t))_{t\geq 0} = ((X_1(t), \ldots, X_d(t))_{t\geq 0}$ describing the evolution over time of the molecular counts trajectory is a Markov process. To this Markov process, we associate a state-space $\mathcal{S}$ defined as the subset of $\mathbb{Z}^d_{\geq 0}$ that is forward invariant and minimal, that is, it is the smallest set $\mathcal{S}$ such that if $X(0) = x_0 \in \mathcal{S}$, then $X_{x_0}(t) \in \mathcal{S}$ for all $t \geq 0$. Such a definition is relevant in the context of the study of the ergodicity properties of networks; for more remarks see [26].

Let $\mathcal{P}(\mathcal{S})$ be the set of all probability distributions on the state-space $\mathcal{S}$ which is endowed with the weak topology. We can then define the probability

$$p_{x_0}(t, x) = \mathbb{P}(X_{x_0}(t) = x) \tag{1}$$

where $x_0, x \in \mathcal{S}$. Defining then $p_{x_0}(t)(A) := \sum_{y\in A} p_{x_0}(t, y)$ where $A \subset \mathcal{S}$, then $p_{x_0}(t)$ can be understood as an element of $\mathcal{S}$ which coincides, actually, with the distribution of the Markov process $(X(s))_{s\geq 0}$ at time t. The evolution of $p_{x_0}(t)$ is governed by the Chemical Master Equation (CME, or Forward Kolmogorov Equation) given by

$$\frac{dp_{x_0}(t, x)}{dt} = \sum_{i=1}^{K} \left[\lambda_k(x - \zeta_k)p_{x_0}(t, x - \zeta_k) - \lambda_k(x)p_{x_0}(t, x)\right]. \tag{2}$$

where $p(0, x) = \delta_{x_0}(x)$ and $\delta(x - x_0)$ is the Kronecker $\delta(\cdot)$ function. In general, the CME is not analytically solvable except in sone particular simple cases; see e.g. [39] and the references therein. This is the reason why numerical solutions are of interest; see e.g. [40, 41] and the references therein. Alternatively, we can write the so-called random time change representation of the system which takes the form

$$X(t) = X(0) + \sum_{i=1}^{K} \zeta_i Y_i \left(\int_0^t \lambda_i(X(s))ds\right) \tag{3}$$

where the Y_i's are independent unit-rate Poisson processes and $X(s) = 0$, $s < 0$. Finally, it is important to define the generator $\mathbb{A}$ of the Markov process representing the reaction network $(\boldsymbol{X}, \mathcal{R})$:

$$\mathbb{A}f(x) = \sum_{i=1}^{K} \lambda_i(x) \left[f(x + \zeta_i) - f(x) \right] \tag{4}$$

for all functions $f : \mathbb{Z}^d_{\geq 0} \to \mathbb{R}$ in the domain of $\mathbb{A}$. For such functions f, Dynkin's formula is valid and we have that

$$\mathbb{E}[f(X(t))] = \mathbb{E}[f(X(s))] + \int_s^t \mathbb{E}[\mathbb{A}f(X(\theta))]d\theta \tag{5}$$

for all $s \leq t$.

2.2 Stochastic reaction networks with delays

Reaction networks with delays are not new and have been studied in the past, both in the deterministic [7, 42–45] and stochastic [10, 11, 21] settings. Let τ_k be the delay of reaction k and decompose the stoichiometric vector ζ_k as $\zeta_k = \zeta_k^r - \zeta_\ell^\ell$ where the $\zeta_k^r \geq 0$ and $\zeta_k^\ell \geq 0$ are the right- and left-stoichiometric vector of the k-th reaction. Then, we can decompose each reaction with positive delay as a sequence of two reactions. The first one happens instantaneously, i.e. when the k-th reaction fires (note that the propensities always depend on the current state $X(t)$), and changes the state value by $x \mapsto x - \zeta_k^\ell$. The second one occurs after τ_k seconds and changes the state value by $x \mapsto x + \zeta_k^r$. This temporal behavior is difficult to capture in the chemical master equation or in the generator without extending the state-space. However, this can be easily incorporated in the random time change representation of the system as it is a temporal characterization of the process [20]. That is, we have that

$$X(t) = X(0) + \sum_{i=1}^{K} \left[-\zeta_i^\ell Y_i \left(\int_0^t \lambda_i(X(s))ds \right) + \zeta_i^r Y_i \left(\int_0^{t-\tau_i} \lambda_i(X(s))ds \right) \right]. \tag{6}$$

where the Y_i's are independent unit-rate Poisson processes and $X(s) = 0$, $s < 0$. Note the coupling of the instantaneous part and the delayed part of the reaction through the same Poisson process. Note, however, that, depending on the application, this may not be the best way to represent a delayed reaction as the species on the left-hand side on the reaction are not necessarily destroyed as in catalytic reactions.

2.3 Types of delays

We discuss here the different types of delays that can be considered. We have to distinguish here two problems: incorporating delays for simulation purpose and incorporating delays for modeling and analysis purposes. The first one is usually easier than the second one as the problem is essentially computational and existing algorithms can sometimes be easily adapted to include delayed reactions. The difficulty with the modeling and the analysis problems

is that the state-space of a delayed reaction network essentially becomes infinite dimensional and this, therefore, make the chemical master equation and the analysis of the state-space more complex.

Deterministic and stochastic constant delays. Probably, the most natural type of delays that comes to mind is the case of deterministic constant delays. As briefly stated in [20], the random time change representation and Gillespie's stochastic simulation algorithm can be adapted to cope with such delays. On the modeling and analysis side, the state-space becomes infinite-dimensional which leads to difficulty for extending the chemical master equation, checking the irreducibility of the state-space and the positive recurrence of the Markov process. The case of stochastic constant delays is analogous. At the beginning of the simulation, the delays are drawn from the distribution and kept constant until the simulation is over. Regarding the modeling and analysis, the problem is the same as for deterministic constant delays.

Deterministic time-varying delays. Existing algorithms should, in principle, be extended to account for deterministic time-varying delays. However, this may lead to a high increase of the computational complexity. Regarding modeling and analysis are more complex than their constant counterparts and, therefore, lead to, at least, the same difficulties.

Stochastic time-varying delays. Stochastic time-varying delays, as we shall see later, are much more natural in this context provided that their distribution is of some particular kind. It has been shown in [23, 24] that bidirectional chains of unimolecular conversion reactions could be substituted, in simulation, by a time-varying stochastic delay, the distribution of which being computed from the reaction rates and the topology of the reaction network to be reduced. This has led to dramatic improvements in terms of simulation time. In this regard, those delays should be easily incorporable in the reaction network model and should also facilitate their analysis. This will be shown to be true when the delay distribution belongs to the class of phase-type distributions; see e.g. [36]. It will be notably demonstrated that the state-space of the delayed reaction network remains finite-dimensional in this case.

2.4 Ergodicity analysis of stochastic reaction networks

The ergodicity of stochastic chemical reaction networks is the important property that the CME has a unique attractive fixed-point, that is, for any initial condition $p(0, x)$, there exists a unique stationary distribution p^* such that $p_{x_0}(t, x) \to p^*(x)$ as $t \to \infty$. Several conditions have been provided in the literature for checking the ergodicity of Markov processes. An interesting approach is based on the so-called Foster-Lyapunov functions introduced in [46]. This approach has been specialized to the case of stochastic reaction networks in [26, 47]. Additional results on the stability of reaction networks have also been provided in [48–50]. In particular, let us consider stochastic reaction networks with second-order mass-action kinetics, then we have the following general result:

Proposition 1 ([46]) *Let us consider the stochastic reaction network* $(\boldsymbol{X}, \mathcal{R})$ *and assume that its state-space is irreducible. Define* $\mathbb{A}$ *be the generator of the underlying Markov process. Let us consider a compact set Assume further that there exists a function* $V : \mathbb{R}^d \mapsto \mathbb{R}$ *such that*

(a) $V(x) > 0$ *for all* $x \notin C$*,*

(b) V *is bounded on* C*, and*

(c) $\mathbb{A}V(x) \leq -cf(x)$ *for some* $c > 0$ *and for all* $x \notin C$

for some compact set $C \subset \mathbb{R}^d$*, some* $c > 0$ *and for some function* $f(x) \geq 1$*. Then, the Markov process is ergodic. If the conditions hold with* $f(x) = V(x)$*, then the Markov process is exponentially ergodic. The function* V *is, moreover, called a Foster-Lyapunov function.*

In the case of stochastic mass-action reaction networks with zeroth- to second-order reactions, it is possible to exploit the positivity of the dynamics of the system and consider a linear positive Foster-Lyapunov function of the form $V(x) = v^T x$ where $v \in \mathbb{R}^d_{>0}$. Before stating the main result, we need few definitions. We define the stoichiometric matrix S of this network as

$$S = \begin{bmatrix} S_0 & S_u & S_b \end{bmatrix} \tag{7}$$

where S_0 corresponds to zeroth order reactions, S_u to first order reactions and S_b to second order reactions. Correspondingly, we define the propensity functions associated with the zeroth- and first-order reactions as $\lambda_0(x) = w_0$ and $\lambda_u(x) = W_u x$.

This leads to the following result:

Theorem 2 ([26]) *Let us consider the stochastic reaction network* $(\boldsymbol{X}, \mathcal{R})$ *and assume that it is mass-action with zeroth- to second-order reactions. Assume that the state-space of the Markov process is irreducible and that there exists a vector* $v \in \mathbb{R}^d_{>0}$ *such that the conditions*

$$v^T S_u W_u < 0 \text{ and } v^T S_b = 0 \tag{8}$$

hold. Then, the reaction network is exponentially ergodic and the stationary distribution is light-tailed.

In the case of unimolecular networks, we have the following corollary:

Corollary 3 ([26]) *Let us consider the stochastic reaction network* $(\boldsymbol{X}, \mathcal{R})$ *and assume that it is mass-action with zeroth- to first-order reactions. Assume further that the matrix* $S_u W_u$ *is nonsingular. Then, the following statements are equivalent:*

(a) There exists a vector $v \in \mathbb{R}^d_{>0}$ *such that* $v^T S_u W_u < 0$*.*

(b) The matrix $S_u W_u$ *is Hurwitz stable.*

(c) The Markov process describing the reaction network is exponentially ergodic and the stationary distribution is light-tailed.

As a consequence, all the moments are bounded and globally converging to their unique stationary value.

Remark 4 *The above results are interesting for different reasons. The first one is that they can be stated in a very simple way and can, therefore, be easily understood by a wide audience. The second one is that the conditions can be numerically verified even for very large systems since it belongs to the tractable class of finite-dimensional linear programming problems for which many powerful algorithms exist; see e.g. [51]. It is important to mention that the number of constraints and variables scale linearly as a function of the number of species, not the number of reactions which can be typically much higher. This was first noticed in [26].*

Remark 5 *In all the above results, the irreducibility of the state-space is tacitly assumed. It seems then important to clarify how this assumption can be checked. First of all, note that given a network topology, the irreducibility of the state-space is a structural property in the sense that if it holds for a given network with certain reaction rates, it will also hold for all possible values of the reaction rates, excluding the zero value. It has been proven that the irreducibility of reaction networks can be checked by solving a linear program and a simple linear algebraic condition [31]. In this regard, proving the ergodicity of reaction networks is possible using simple algebraic and computational methods. Note, however, that the conditions are in general sufficient only, but it has been emphasized in [26] that they have been able to successfully prove the ergodicity of several typical reaction networks considered in the literature.*

2.5 Antithetic integral control of stochastic reaction networks

The antithetic integral controller has been introduced in [38] with the aim of developing an integral control theory for stochastic chemical cellular processes. Integral control is a cornerstone of control theory and engineering as it allows to steer the output of a given system towards a desired constant set-point and to regulate this output around this value despite the presence of constant disturbances acting on the system. The idea behind the antithetic integral control is that it needs to be amenable to chemical reactions in order to envisage the possibility of implement it in-vivo [52] This controller takes the form of a stochastic reaction network $(\boldsymbol{X} \cup \mathcal{Z}, \mathcal{R}_{\text{AIC}})$ with species $\mathcal{Z} := \{\boldsymbol{Z_1}, \boldsymbol{Z_2}\}$ and reactions

$$\boldsymbol{Z_1} \xrightarrow{k} \boldsymbol{Z_1} + \boldsymbol{X_1}, \boldsymbol{X_\ell} \xrightarrow{\theta} \boldsymbol{X_\ell} + \boldsymbol{Z_2}, \emptyset \xrightarrow{\mu} \boldsymbol{Z_1}, \boldsymbol{Z_2} + \boldsymbol{Z_1} \xrightarrow{\eta} \emptyset \tag{9}$$

where $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are the actuating and the sensing species, respectively. The species $\boldsymbol{X_\ell}$ is the measured/controlled species we would like to control by acting on the production rate of the actuated species $\boldsymbol{X_1}$. The first reaction is referred to as the actuation reaction since it catalytically produces one molecule of the actuated species with a rate proportional to the actuating species. Symmetrically, the second reaction is the sensing reaction since it catalytically produces one molecule of the sensing species with a rate proportional to the measured species. The third reaction is the reference reaction since it sets part of the

set-point for the stationary mean of the controlled species population. Finally, the last reaction is the comparison as it compares the populations of the controller species and acts as a nonlinear subtraction operator while, at the same time, closing the loop. Without this reaction the interconnected network would not be in closed loop since the populations of the controller species would be completely uncorrelated.

When the closed-loop network $(\boldsymbol{X} \cup \mathcal{Z}, \mathcal{R} \cup \mathcal{R}_{\text{AIC}})$ is ergodic, a quick inspection at the first-order moment dynamics allow us to state that

$$\mathbb{E}[X_\ell(t)] \to \mathbb{E}_\pi[X_\ell] := \mu/\theta \text{ as } t \to \infty \tag{10}$$

where $\mathbb{E}_\pi$ denotes the expectation operator at stationarity. In this regard, the antithetic integral controller allows to steer to mean population of the controlled species to a desired set-point. It also able to achieve perfect adaptation for the closed-loop dynamics provided that the set-point is achievable; i.e. there must exist a positive steady-state for the closed-loop dynamics for which we have $\mathbb{E}_\pi[X_\ell] = \mu/\theta$.

In the case of unimolecular networks with mass-action kinetics, then we have the following result:

Proposition 6 ([38]) *Let us consider the stochastic reaction network $(\boldsymbol{X}, \mathcal{R})$ and assume that it only contains zeroth- and first-order reactions with mass-action kinetics. Let us further assume that the state-space of the underlying reaction network is irreducible and that there exist vectors $v \in \mathbb{R}^d_{>0}$ and $w \in \mathbb{R}^d_{\geq 0}$, $w_1, w_\ell > 0$,*

(a) $v^T S_u W_u < 0$,

(b) $w^T S_u W_u + e_\ell^T = 0$, *and*

(c) $\dfrac{\mu}{\theta} > \dfrac{v^T S_0 \lambda_0}{v_\ell}$

where $\{e_i\}$ is the standard basis of the Euclidian space and λ_0 is the propensity vector associated with zeroth-order reactions. Then, the closed-loop network $(\boldsymbol{X} \cup \mathcal{Z}, \mathcal{R} \cup \mathcal{R}_{\text{AIC}})$ is ergodic and we have that

$$\mathbb{E}[X_\ell(t)] \to \mu/\theta \text{ as } t \to \infty. \tag{11}$$

As for the ergodicity tests, the conditions stated in the above theorem can be solved using standard linear programming techniques. Moreover, it is interesting to note that the conditions are independent of the gain k and the annihilation parameter η of the controller. This is quite surprising as it is well-known that setting the gain of an integral controller too high results into a destabilization of the closed-loop dynamics unless the system to be controlled is strictly passive – a quite strong property.

Interestingly, when there is no zeroth-order reactions, the above result can be reformulated in terms of control theoretic concepts:

Proposition 7 ([38]) *Let us consider the stochastic reaction network* $(\boldsymbol{X}, \mathcal{R})$ *and assume that it only contains first-order reactions with mass-action kinetics. Let us further assume that the state-space of the underlying reaction network is irreducible. Then, the following statements are equivalent:*

(a) The following conditions hold:

(i) the matrix $S_u W_u$ *is Hurwitz stable, and*

(ii) the system $(S_u W_u, e_1, e_\ell^T)$ *is output controllable.*

(b) The closed-loop network $(\boldsymbol{X} \cup \mathcal{Z}, \mathcal{R} \cup \mathcal{R}_{\text{AIC}})$ *is ergodic.*

As a result, we have that

$$\mathbb{E}[X_\ell(t)] \to \mu/\theta \text{ as } t \to \infty. \tag{12}$$

It is worth mentioning that the antithetic integral controller can be used to control more complex networks including bimolecular reactions or non-mass-action kinetics such as Michaelis-Menten kinetics [38]. However, the theoretical results do not cover most of the interesting cases because of the potential local loss of the output-controllability condition. Simulation results tend to suggest that this controller is functioning properly even when the output-controllability property is not a global property, as is often the case in nonlinear systems.

3 Phase-type distributed delays and their reaction network implementation

The aim of this section is to convince the reader that delays obeying phase-type distributions naturally arise in stochastic reaction networks. In this regard, they are the obvious choice to work with. We first recall some theoretical basics on phase-type distributions and provide some examples in order to demonstrate their richness in terms of behavior. In fact, those distributions are known to be dense in the set of all probability distributions, which means that we can approximate arbitrarily closely any distribution, including heavy-tailed distributions. We then provide a complete characterization of such distributions in terms of simple algebraic conditions. Finally, under the existence assumption of a Markov process describing the considered probability distribution, we propose a simple procedure to construct a minimal unimolecular reaction network encoding this distribution. By minimal, it is meant that it contains the smallest number of molecular species and reactions that any other network encoding the same distribution.

3.1 Preliminaries on phase-type distributions

A phase-type distribution is a combination of mixtures and convolutions of exponential distributions. It is obtained by forming a system of interrelated Poisson processes placed in

series (also known as phases). It can be represented by a random variable describing the time until absorption of a Markov process with one absorbing state and with initial condition α, each state of the Markov process representing a phase of the overall process. The probability density function of the phase-type distribution $\mathrm{PH}(\alpha, H)$ is given by

$$f(\tau) = \alpha e^{H\tau} H_0, \ \tau \geq 0 \tag{13}$$

where $\alpha \in \mathbb{R}_{\geq 0}^{1\times n}$, $||\alpha||_1 = 1$, H is a Hurwitz stable Metzler matrix and $H_0 = -H\mathbb{1}_n$. The cumulative distribution is given by

$$F(\tau) = 1 - \alpha e^{H\tau}\mathbb{1}_n, \ \tau \geq 0. \tag{14}$$

The Laplace transform of the distribution is given by

$$\widehat{f}(s) = \alpha(sI - H)^{-1} H_0 \tag{15}$$

which verifies $\widehat{f}(0) = 1$. Let $X \sim \mathrm{PH}(\alpha, H)$, then all the moments of this random variable are given by

$$\mathbb{E}[X^\ell] = (-1)^\ell \ell! \alpha H^{-\ell} H_0. \tag{16}$$

The evolution of the probability distribution of the corresponding Markov process is described by the forward Kolmogorov equation

$$\dot{p}(t) = p(t) \begin{bmatrix} 0 & 0_{1\times n} \\ H^0 & H \end{bmatrix} \text{ with } p(0) = \begin{bmatrix} 0 & \alpha \end{bmatrix}. \tag{17}$$

Solving for this differential equation yields

$$p(t) = \begin{bmatrix} 0 & \alpha \end{bmatrix} \exp\left(\begin{bmatrix} 0 & 0_{1\times n} \\ H^0 & H \end{bmatrix} t \right) \begin{bmatrix} 1 - \alpha e^{Ht}\mathbb{1} & \alpha e^{Ht} \end{bmatrix} \tag{18}$$

where we can recognize the expression of the cumulative distribution.

3.2 Examples

We give here few examples of phase-type distributions.

Hypoexponential and Erlang distributions. Hypoexponential distributions consist of the convolution of a finite number of exponential distributions with possibly different various rates. When all the rates are equal, the hypoexponential distribution reduces to the Erlang distribution. Examples of Erlang distributions are given in Figure 1 where we can observe that this distribution is very rich as it can take various forms. In the case of a hypoexponential distribution with four phases and with parameters $\lambda_1, \ldots, \lambda_4 > 0$, the matrix H is given by

$$H = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 \\ 0 & 0 & 0 & -\lambda_4 \end{bmatrix}. \tag{19}$$

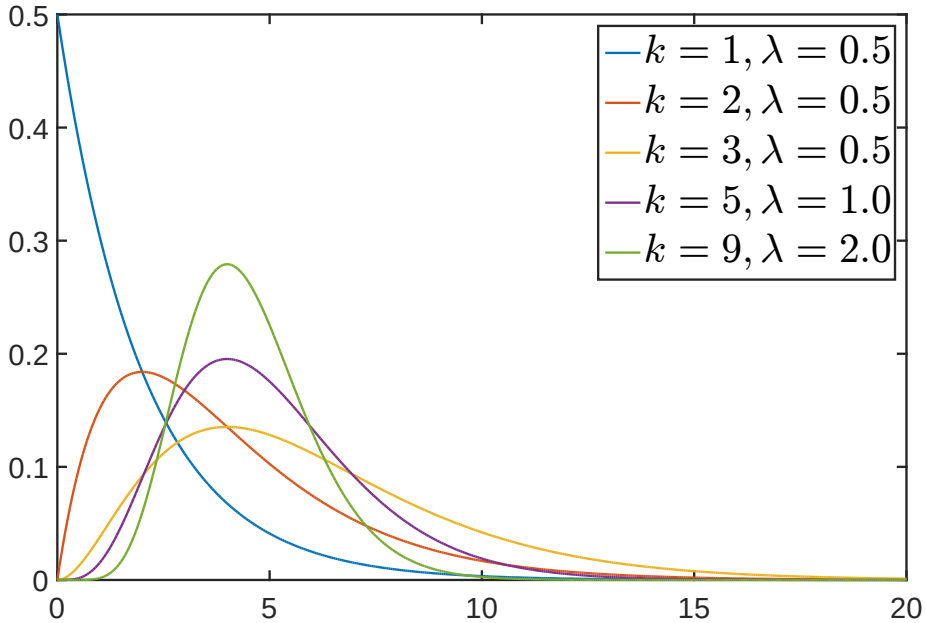

Figure 1: Examples of Erlang distributions

Typical hypoexponential distributions are depicted in Figure 2 for randomly chosen parameters $\lambda_1, \ldots, \lambda_4 > 0$.

Hyperexponential and Hyper-Erlang distributions. Hyperexponential distributions consist of the mixture of exponential distributions, that is, the density function of a hyperexponential distribution is a convex combination of density functions of exponentially distributed random variables. Analogously, the Hyper-Erlang distribution has a density function consisting of a convex combination of density functions of Erlang random variables. For instance, let us consider two Erlang distributions. The parameter and the number of stages of the first one are 5 and 20, respectively. The second one has 5 and 80 as parameter and number of stages. The density function of the considered Hyper-Erlang depicted in Figure 3 consists of the average of the density functions associated with the aforementioned Erlang distributions.

3.3 Approximation of a constant delay

In the deterministic setting, constant delays can be approximated arbitrarily closed by sequence of filters such as low-pass filters [53, 54] or all-pass filters such as lattice networks or Padé approximants [53, 55, 56]. It is notably very well-known that the constant delay operator $\nabla_{\bar{\tau}}$ with delay $\bar{\tau} > 0$ having $e^{-\bar{\tau}s}$ as transfer function can be approximated arbitrarily well by a sequence of N low-pass filters placed in series with overall transfer function given

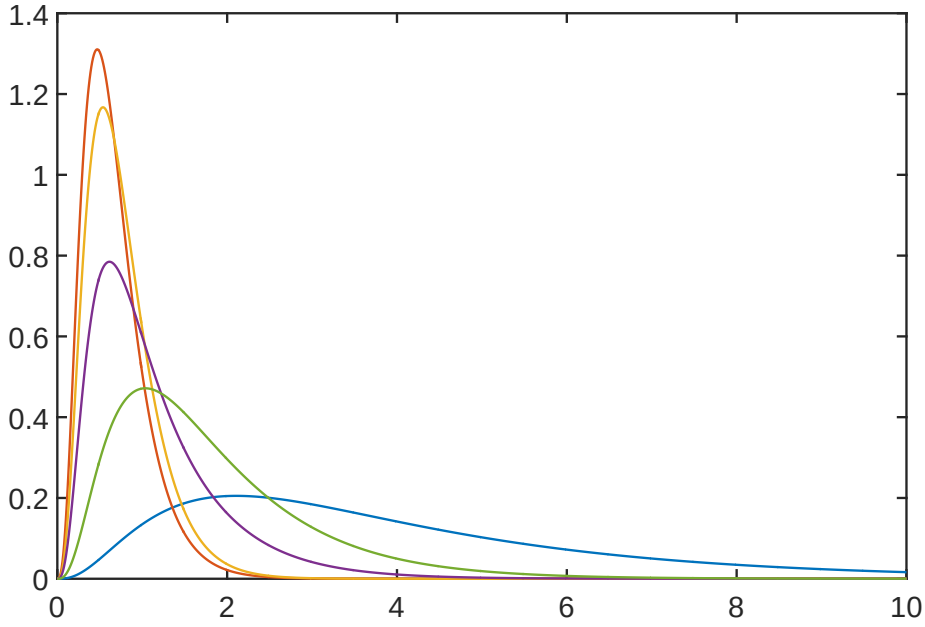

Figure 2: Examples of hypoexponential distributions

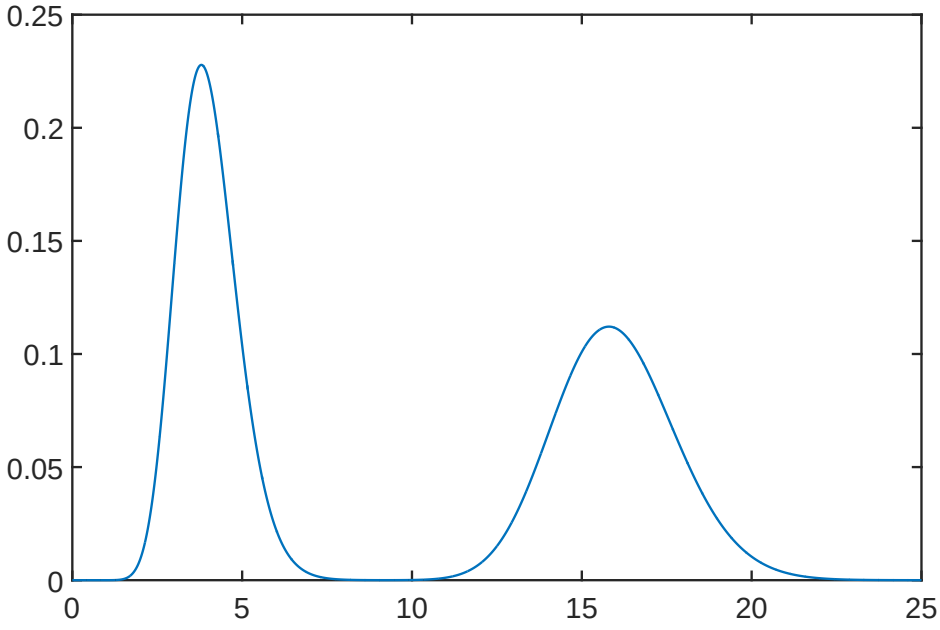

Figure 3: Example of a bimodal Hyper-Erlang distribution.

by

$$H_N(s) := \frac{1}{\left(1+\frac{\bar{\tau}s}{N}\right)^N}, N \in \mathbb{Z}_{>0}. \tag{20}$$

Note, moreover, that $e^{-\bar{\tau}s}$ is the Laplace transform of the Dirac distribution $\delta(t-\bar{\tau})$ centered around $\bar{\tau}$.

The above statement is formally stated in the following standard result which is recalled for completeness:

Proposition 8 ([53, 54]) *We have that $H_N(s) \to e^{-s\bar{\tau}}$, for all $s \in \mathbb{C}$, as $N \to \infty$.*

Proof : Clearly, $H_N(s)$ can be rewritten as $H_N(s) = \exp(\log(H_N(s)))$. Hence,

$$\begin{aligned} H_N(s) &= \exp(\log(H_N(s))) \\ &= \exp(-N\log(1+\tau s/N)) \\ &\simeq \exp(-\tau s)) \text{ since for large } N \text{ we have } \log(1+\tau s/N) \simeq \tau s/N \end{aligned} \tag{21}$$

and hence we have that $H_N(s) \to e^{-s\bar{\tau}}$, for all $s \in \mathbb{C}$, as $N \to \infty$. This completes the proof. $\diamondsuit$

Interestingly, a similar result exists in the stochastic setting. Indeed, a constant delay $\bar{\tau}$ can be expressed as the limit of an Erlang random variable. This is stated in the following result:

Proposition 9 *Let us then consider a random variable τ_N following an Erlang distribution with shape N (i.e. number of phases) and rate $N/\bar{\tau}$. The corresponding density function is given by*

$$f_N(x) = \frac{N^N x^{N-1} e^{-Nx/\bar{\tau}}}{\bar{\tau}^N (N-1)!}. \tag{22}$$

Then, we have that $\tau_N \to \bar{\tau}$ in distribution as $N \to \infty$.

Proof : Instead of proving that $f_N(x) \to \delta(x-\bar{\tau})$ as $N \to \infty$, we propose a simpler alternative approach based on the mean and variance of the random variable τ_N. From the moments expressions in (16), it is immediate to see that the mean of τ_N is given by

$$\mathbb{E}[\tau_N] = \bar{\tau} \tag{23}$$

and its variance by

$$V(\tau_N) = \bar{\tau}^2/N. \tag{24}$$

Hence, we have that $\mathbb{E}[\tau_N] \to \bar{\tau}$ and $V(\tau_N) \to 0$ as $N \to \infty$, which means that the density function converges to the shifted Dirac delta $\delta(x-\bar{\tau})$, which is the distribution of a deterministic constant. This proves the result. $\diamondsuit$

The above result shows that, in the limit, the Erlang distribution tends to a Dirac with mass localized at $\bar{\tau}$. Figure 4 depicts different Erlang distribution where we can see that as N increases the shape of the Erlang distribution gets closer to the Dirac distribution.

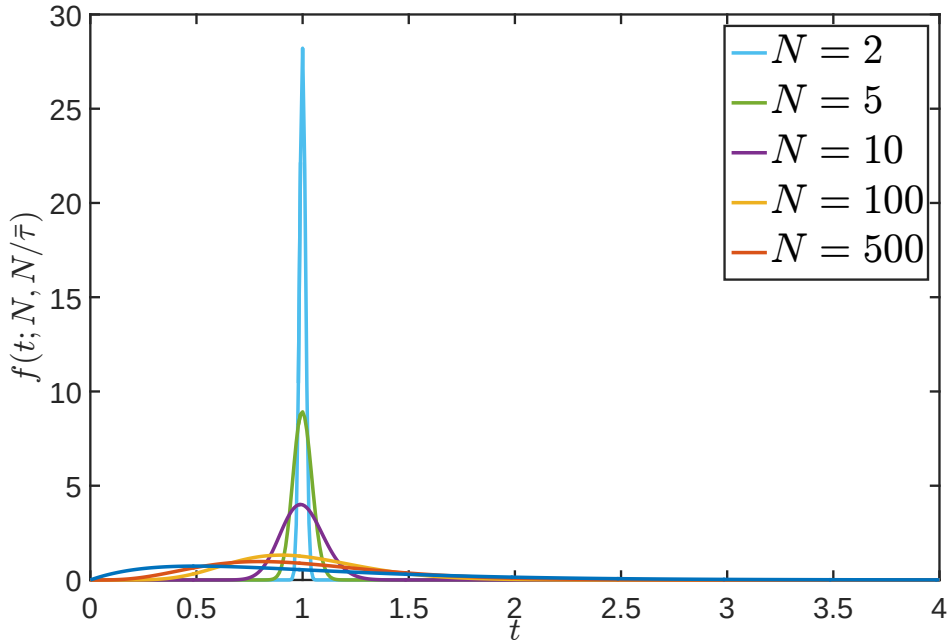

Figure 4: Erlang distributions with rate $N/\bar{\tau}$ and shape N approximating the Dirac distribution.

3.4 Reaction network implementation of phase-type delays

We now address two problems. The first one is starting from Hurwitz stable Metzler matrix $H \in \mathbb{H}^{n\times n}$, does this matrix can be used to represent a phase-type distribution. And, if so, how can we construct a conservative reaction network that encodes such distribution. Additional properties of the reaction network are also studied.

Existence of a Markov process. The following result fully characterizes whether a given matrix $H \in \mathbb{R}^{m\times m}$ can be used to model a delay which phase-type distribution:

Proposition 10 *Let us consider a matrix $H \in \mathbb{R}^{m\times m}$. The matrix H generates the transition of a Markov process if and only if*

1. *it is Metzler and Hurwitz stable,*
2. *$H\mathbb{1}_m \leq 0$, and*
3. *at least one entry of $H\mathbb{1}_n$ is nonzero; i.e. $H\mathbb{1}_m \neq 0$.*

When these conditions are verified, the matrix H is said to be ***admissible****.*

Proof : The matrix H generates the transition of a Markov process if and only if the matrix

$$\begin{bmatrix} 0 & 0 \\ -H\mathbb{1} & H \end{bmatrix} \tag{25}$$

is a Q-matrix. For this to hold, we need that the rows sum to zero, which is the case. We also need the matrix to be Metzler, which is the case if and only if $H\mathbb{1}_m \leq 0$ holds. Finally, $H\mathbb{1}_m$ cannot be zero, otherwise there would be no absorbing state. $\diamondsuit$

The above conditions are very easy to check and can also be used to characterize all the matrices H that can be considered.

Construction of the reaction network. We now address the problem of constructing a stochastic reaction network encoding a given phase-type distributed delay. Since the matrix H is of dimension n, the network needs to have at least m phases and, hence, m molecular species. In fact, we need exactly m species. We denote those species by $\boldsymbol{D_1}, \ldots, \boldsymbol{D_m}$. In addition, the reaction network must be satisfy the following property

- The delay line must be conservative in the sense that nothing is lost or created inside the queue.
- The state-space of the reaction network describing the delay line need to be irreducible.

The first constraint is easily fulfilled by considering exclusively conversion reactions. Catalytic or degradation reactions cannot be used as they do not preserve mass. Multimolecular reactions cannot be used since their propensity is nonlinear, while we need linear propensities to appropriately represent the matrix H. The second constraint may seem contradictory with the fact that the Markov process describing the delay distribution has an absorbing state. Recall that the delay is the time-to-absorbtion of this Markov process. However, what we are requiring here is that the state-space $\mathbb{Z}^m_{\geq 0}$ of the stochastic reaction network be irreducible. The motivation behind this constraint is that we would like to obtain conditions for the ergodicity of delayed stochastic reaction networks.

The following result states how to construct the minimal stochastic reaction reaction network implementing a given phase-type distributed delay:

Proposition 11 *Assume that $H = [h_{ij}]_{i,j=1,\ldots,m}$ satisfies the conditions of Proposition 10 and let $\alpha \in \mathbb{R}^{1\times n}$ be a probability row vector. Then, the delay $\tau \sim \mathrm{PH}(\alpha, H)$ is exactly represented by the stochastic reaction network*

$$\emptyset \xrightarrow{\alpha e_i} \boldsymbol{D_i}, \qquad \boldsymbol{D_i} \xrightarrow{h_{ij}} \boldsymbol{D_j}, \qquad \boldsymbol{D_i} \xrightarrow{-e_i^T H \mathbb{1}} \emptyset, \tag{26}$$

where $i, j = 1, \ldots, m$, $i \neq j$. Moreover, this network is minimal (minimal number of species and reactions) and its state-space is irreducible.

Proof : The queue is conservative since it only consists of conversion reactions. The irreducibility is also immediate since we can see that any state can be reached from any other state through a sequence of reactions having a positive propensity. The minimality of the species comes from the fact that each species correspond to one phase. Since, we have m phases, we need at least m species. Finally, the minimality of the reactions comes from the fact that we have exactly one reaction for every nonzero off-diagonal entry of the matrix H.

In this regard, every reaction is indispensable. This proves the minimality of the number of reactions. This proves the result. ♢

The above result shows that any phase-type distributed delay can be represented as a simple unimolecular chemical reaction network. This is very convenient since it is known from [26, 28, 29] that unimolecular stochastic reaction networks are usually well-behaved and easy to analyze using standard linear algebra tools.

Example. Let us consider a delay that follows the Erlang distribution with rate $\lambda > 0$ and shape m. In this case, we have that

$$H = \begin{bmatrix} -\lambda & \lambda & 0 & \dots & 0 \\ 0 & -\lambda & \lambda & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & -\lambda \end{bmatrix}, H_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ \lambda \end{bmatrix}, \alpha = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}^T \tag{27}$$

Then, the corresponding reaction network is given by

$$\emptyset \xrightarrow{1} \boldsymbol{D_1}, \quad \boldsymbol{D_i} \xrightarrow{\lambda} \boldsymbol{D_{i+1}}, \quad \boldsymbol{D_m} \xrightarrow{\lambda} \emptyset, i = 1, \dots, m-1. \tag{28}$$

4 Stochastic reaction networks with phase-type distributed delays

We know now that for any phase-type distributed delay corresponds a minimal unimolecular stochastic reaction network. We are then in position to define delayed reactions in reaction networks. Delayed reactions take the form

$$f(\boldsymbol{X}) \xrightarrow{k, \tau(H,\alpha)} g(\boldsymbol{X}) \tag{29}$$

where $f(\boldsymbol{X})$ and $g(\boldsymbol{X})$ denote any combination of reactants and products, respectively.

The first parameter of the reaction is the *reaction rate*, assumed to be nonnegative, whereas the second one is the *delay* where $\tau(H, \alpha)$ is a shorthand for $\tau \sim \mathrm{PH}(H, \alpha)$.

Assume that we have a reaction with delay of the form

$$f(\boldsymbol{X}) \xrightarrow{k, \tau(H,\alpha)} g(\boldsymbol{X}) \tag{30}$$

where $f(\boldsymbol{X})$ and $g(\boldsymbol{X})$ denote any type of constructs made from the species in $\boldsymbol{X}$. Assume that the phase-type distribution has m phases. We can then apply the construction procedure of Section 3.4 to get the delay the delay line. However, we now interconnect it with the input

of the delay lines with the reactants and the output of the delay line with the products of the reaction. This yields the following possible realization

$$f(\boldsymbol{X}) \xrightarrow{k,\alpha e_i} \boldsymbol{D_i} \quad \text{and} \quad \boldsymbol{D_i} \xrightarrow{-e_i^T S \mathbb{1}_n} g(\boldsymbol{X}) \tag{31}$$

for all $i = 1, \ldots, m$. It is important to note that the first reaction is destructive in the sense that the species on the left-hand side of the reaction are destroyed when the input reaction fires. Also, the absorbed molecular species will not be available while being in the line. This may not be compatible with all reactions, in particular, catalytic reactions. In this case, the following realization should be considered

$$f(\boldsymbol{X}) \xrightarrow{k\alpha e_i} f(\boldsymbol{X}) + \boldsymbol{D_i} \quad \text{and} \quad \boldsymbol{D_i} \xrightarrow{-e_i^T S \mathbb{1}_n} g(\boldsymbol{X}). \tag{32}$$

In this case, we can easily see that the input reaction is non destructive and that $f(\boldsymbol{X})$ will be available for reactions while there will be some molecules in line. The choice of the correct input and output reactions highly depend on the context.

Example. Let us consider for instance the catalytic reaction

$$\boldsymbol{X_1} \xrightarrow{k,\tau(H,\alpha)} \boldsymbol{X_1} + \boldsymbol{X_2} \tag{33}$$

where the delay is assumed to be Erlang with rate λ and shape m. This network can have different behaviors. The first one is that when the reaction fires, only the product production is delayed while the molecule of $\boldsymbol{X_1}$ involved can still participate to reactions. In this case, we have the following reaction network implementation:

$$\boldsymbol{X_1} \xrightarrow{k} \boldsymbol{X_1} + \boldsymbol{D_1}, \quad \boldsymbol{D_i} \xrightarrow{\lambda} \boldsymbol{D_{i+1}}, \quad \boldsymbol{D_m} \xrightarrow{\lambda} \boldsymbol{X_2}, i = 1, \ldots, m-1. \tag{34}$$

The second one is when the involved molecule of $\boldsymbol{X_1}$ is monopolized by the delayed reaction and cannot participate to other reactions. In this case, we have

$$\boldsymbol{X_1} \xrightarrow{k} \boldsymbol{D_1}, \quad \boldsymbol{D_i} \xrightarrow{\lambda} \boldsymbol{D_{i+1}}, \quad \boldsymbol{D_m} \xrightarrow{\lambda} \boldsymbol{X_1} + \boldsymbol{X_2}, i = 1, \ldots, m-1. \tag{35}$$

5 Ergodicity analysis of delayed reaction networks with phase-type distributed delays

5.1 State-space irreducibility

The following result states the condition under which the delayed reaction network with phase-type distributed delays has an irreducible state-space:

Proposition 12 *Let us consider a delayed reaction network with phase-type distributed delays. Then, it is irreducible if and only if the delay-free network network is irreducible.*

Proof : Clearly, if the delayed network is irreducible then the delay-free must also be irreducible. To prove the converse, just observe that the delay-lines have an irreducible state by construction, hence the irreducibility only depends on the reaction network with the delays set to 0. This proves the result. $\diamondsuit$

5.2 Ergodicity of unimolecular networks with delays

Let us consider here a unimolecular network with d species, K reactions and n delays. Each delay is phase-type distributed, i.e. $\tau_k \sim \mathrm{PH}(H_i, \alpha_i)$, $i = 1, \ldots, K$, for some admissible matrix $H_i \in \mathbb{R}^{d_i \times d_i}$ and some probability row vector $\alpha_i \in \mathbb{R}^{d_i}_{\geq 0}$, $||\alpha_i||_1 = 1$. Let us also define the stoichiometric matrix S of the delayed reaction network as

$$H = \left[\begin{array}{c|c|c|c} S_x & S_{\mathrm{in},x} & S_{\mathrm{out},x} & 0 \\ \hline 0 & S_{\mathrm{in},d} & S_{\mathrm{out},d} & S_d \end{array}\right] \tag{36}$$

where S_x is the stoichiometric matrix associated with the reactions that do not involve any delay chemical species and, conversely, S_d is the stoichiometric matrix associated with reactions that only involve delay species. The stoichiometric matrix $S_{\mathrm{in}} = \begin{bmatrix} S_{\mathrm{in},x} \\ S_{\mathrm{in},d} \end{bmatrix}$ corresponds to conversion reactions entering the delay lines whereas $S_{\mathrm{out}} = \begin{bmatrix} S_{\mathrm{out},x} \\ S_{\mathrm{out},d} \end{bmatrix}$ corresponds to conversion reactions leaving the delay lines.

The propensity functions are defined as

$$\begin{array}{rcl} \lambda_x(x) & = & W_x x + b_x \\ \lambda_{\mathrm{in}}(x) & = & W_{\mathrm{in}} x + b_{\mathrm{in}} \\ \lambda_{\mathrm{out}}(d) & = & W_{\mathrm{out}} d \\ \lambda_d(d) & = & W_d d \end{array} \tag{37}$$

Let us also define the following matrices

$$\begin{array}{rcl} A & := & S_x W_x + S_{\mathrm{in},x} W_{\mathrm{in}} \\ B & := & S_{\mathrm{out},x} W_{\mathrm{out}} \\ C & := & S_{\mathrm{in},d} W_{\mathrm{in}} \\ D & := & S_{\mathrm{out},d} W_{\mathrm{out}} + S_d W_d \end{array} \tag{38}$$

Calculations show that

$$\begin{array}{rcl} W_{\mathrm{out}} & := & -\operatorname*{diag}_i(\mathbb{1}^T H_i^T) \\ C & := & \begin{bmatrix} \alpha_1^T q_1^T \\ \vdots \\ \alpha_N q_N^T \end{bmatrix} \\ D & = & H^T := \operatorname*{diag}_i(H_i)^T \end{array} \tag{39}$$

where $q_j = r_j e_{\sigma(j)}$ where r_j is the reaction rate of the delayed reaction associated with the delay line j and $e_{\sigma(j)}$ is the vector of zeros except at the entry $\sigma(j)$ where it is one. Finally, $\sigma(j)$ is the mapping $\sigma : \{1, \ldots, N\} \mapsto \{1, \ldots, d\}$ where $\sigma(j) = i$ if the species $\boldsymbol{X_i}$ enters the delay line j.

Interestingly, the stoichiometric matrix associated with the delay-free reaction network is given by

$$S_0 := \begin{bmatrix} S_x & S_{\text{in},x} + S_{\text{out},x} \end{bmatrix} \tag{40}$$

with the propensity function

$$\lambda_0(x) := \begin{bmatrix} W_x x + b_x \\ W_{\text{in}} x + b_{\text{in}} \end{bmatrix} \tag{41}$$

which leads to the characteristic matrix

$$A_0 = S_x W_x + (S_{\text{in},x} + S_{\text{out},x}) W_{\text{in}}. \tag{42}$$

We then have the following result:

Theorem 13 *Assume that the state-space of the delay-free network is irreducible and that the matrix A is invertible. Then, the following statements are equivalent:*

1. *The delayed network is exponentially ergodic.*
2. *The matrix*
$$\mathcal{A} := \left[\begin{array}{c|c} A_0 & B \\ \hline C & H^T \end{array}\right]$$
is Hurwitz stable.
3. *The matrix $A + \sum_i S^i_{\text{out},x} q_i^T$ is Hurwitz stable where $S^i_{\text{out},x}$ denotes the i-th column of $S_{\text{out},x}$.*
4. *The delay-free network (i.e. the network with all the delays set to zero) is ergodic.*

Proof : Let $\tilde{\mathbb{A}}$ be the generator of Markov process representing the delayed reaction network and define the linear form $V(x, \delta) := v^T x + \sum_{i=1}^n v_{d,i}^T \delta_i$ defined for $v \in \mathbb{R}^d_{>0}$ and $v_{d,i} \in \mathbb{R}^{d_i}_{>0}$, $i = 1, \ldots, n$. Since the state-space of the delay-free reaction network is irreducible, then so is the state-space of the delayed reaction network.

So, the Markov process is exponentially ergodic if and only if there exist $c_1 \geq 0$ and $c_2 > 0$ such that $\tilde{\mathbb{A}} V(x, \delta) \leq c_1 - c_2 V(x, \delta)$ for all $(x, \delta_1, \ldots, \delta_n) \in \mathbb{N}_0^{d+d_1+\ldots+d_n}$. This can be reformulated as

$$\tilde{\mathbb{A}} \left(v^T x + \sum_{i=1}^N v_{d,i}^T \delta_i \right) = \left[\begin{array}{c} v \\ \hline v_{d,1} \\ \vdots \\ v_{d,n} \end{array}\right]^T \left(\left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] \left[\begin{array}{c} x \\ \hline \delta_1 \\ \vdots \\ \delta_n \end{array}\right] + \left[\begin{array}{c} S_x b_x \\ \hline S_{\text{in},d} b_{\text{in}} \end{array}\right] \right) \tag{43}$$

where we have used the fact that $S_{\text{in},x} b_{\text{in}} = 0$ by construction. So, there exists some $c_1 \geq 0$ and $c_2 > 0$ such that $\tilde{\mathbb{A}} V(x, \delta) \leq c_1 - c_2 V(x, \delta)$ for all $(x, \delta_1, \ldots, \delta_n) \in \mathbb{Z}_{\geq 0}^{d+d_1+\ldots+d_n}$ if and only if $\mathcal{A}$ is Hurwitz stable where we have used the invertibility assumption on the matrix A. This proves the equivalence between the two first statement. To prove the equivalence with the third statement, we use the following result on Metzler matrices.

Lemma 14 *Let us consider a Metzler matrix $M = [M_{i,j}]_{i,j=1}^2$ where both M_{11} and M_{22} are square. Then, the matrix M is Hurwitz stable if and only if the matrix $M_{11} - M_{12} M_{22}^{-1} M_{21}$ and M_{22} is invertible.*

Lemma 15 *Let us consider a Metzler matrix $M = [M_{i,j}]_{i,j=1}^2$ where both M_{11} and M_{22} are square. If the matrix M is Hurwitz stable, then the matrices M_{11} and M_{22} are Hurwitz stable.*

In this regard, the Hurwitz stability of $\mathcal{A}$ is equivalent to the Hurwitz stability of $A - BH^{-T}C$ (which implies the Hurwitz stability of A since $A - BH^{-T}C \geq A$, where the comparison sign is componentwise). Since A is invertible stable by assumption, then this is equivalent to the Hurwitz stability of $A - BH^{-T}C$. Using (38)-(39), we obtain the condition in statement (c). Noticing, finally, that $A + \sum_i S_{\text{out},x}^i q_i^T$ is nothing else but the characteristic matrix A_0 of the delayed network with all the delays set to 0; i.e. the delay-free network. This proves the result. $\diamondsuit$

What the above result says is that the shape of the delay distributions is unimportant as long as they are phase-type distributions. In this regard, one can consider *almost constant deterministic delays* in the sense that we can pick an arbitrarily large N in the Erlang distribution in Proposition 9 while still preserving the ergodicity of the network provided that the delay free network is ergodic.

It is also interesting to discuss the invertibility assumption of the matrix A. First of all, this assumption is made to ensure the equivalence between the statements. Otherwise, the condition in the second statement is sufficient only. Moreover, the matrix A is singular whenever a linear combinations of some species molecular counts does not change value over time. An typical example consists of the reactions

$$\boldsymbol{X_1} \xrightarrow{k_1} \boldsymbol{X_2}, \boldsymbol{X_2} \xrightarrow{k_1} \boldsymbol{X_1}. \tag{44}$$

In this case, we have that $X_1(t) + X_2(t) = X_1(0) + X_2(0)$ for all $t \geq 0$. The matrix $A = A_0$, in this case, is given by $\begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix}$, which is obviously singular. However, the process is exponentially ergodic since the Markov process has a finite irreducible state-space. So, when the matrix A is singular, we can sometimes remove the species that are always finite and modifying few reactions involving them to obtain a reduced network having an invertible matrix A. Let us add the following reactions to the above network

$$\boldsymbol{X_2} \xrightarrow{k_3} \boldsymbol{X_2} + \boldsymbol{X_3}, \boldsymbol{X_3} \xrightarrow{k_4} \emptyset. \tag{45}$$

The A matrix for this system is given by

$$\begin{bmatrix} -k_1 & k_2 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \end{bmatrix} \tag{46}$$

which is singular. We now remove the two reactions first reactions and assume that the reaction $\boldsymbol{X_2} \xrightarrow{k_3} \boldsymbol{X_2} + \boldsymbol{X_3}$, which can be rewritten here as $\emptyset \xrightarrow{k_3 X_2} \boldsymbol{X_3}$, always fires at maximal rate, in the sense, that we rewrite this reaction

$$\emptyset \xrightarrow{k_3 \mathbb{1}^T X(0)} \boldsymbol{X_3}. \tag{47}$$

The matrix A of this network is simply $A = -k_4$, which is invertible. This procedure can always be performed when the finitely populated species are involved in production reactions. When those species are involved in destructive reactions like

$$\emptyset \xrightarrow{k_3} \boldsymbol{X_3}, \boldsymbol{X_2} + \boldsymbol{X_3} \xrightarrow{k_4} \emptyset, \tag{48}$$

this procedure does not exist since we would have to replace $k_4 X_2 X_3$ by 0, which is the minimum value for $X_2(t)$. Note also that it is unclear whether this network is ergodic, too.

5.3 Ergodicity of bimolecular networks with delays

5.3.1 No bimolecular reactions enters the queue

Let S_b be the restriction of the stoichiometric matrix associated with the bimolecular reactions. Since no bimolecular reaction enters any delay line, then this matrix can be written as

$$S_b = \left[\begin{array}{c} S_{b,x} \\ \hline 0 \end{array}\right]. \tag{49}$$

We then have the following result:

Theorem 16 *Assume that the state-space of the bimolecular reaction network is irreducible. Then, the following statements are equivalent:*

1. *There exist vectors $v \in \mathbb{R}^d_{>0}$, $v_{d,i} \in \mathbb{R}^{d_i}_{>0}$, $i = 1, \ldots, n$ such that the conditions $\left[v^T \;\middle|\; v_d^T\right] S_b = 0$ and*

$$\left[\begin{array}{c} v \\ \hline v_{d,1} \\ \vdots \\ v_{d,N} \end{array}\right]^T \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] < 0$$

2. *There exist a vector $v \in \mathbb{R}^d_{>0}$ satisfying $v^T S_b^1 = 0$ and $v^T \left(A + \sum_i S^i_{\text{out},x} q_i^T \right) < 0$.*

Moreover, when one of above statements holds, the Markov process describing the delayed reaction network $(\boldsymbol{X}, \mathcal{R})$ *is exponentially ergodic and the stationary distribution is light-tailed.*

Proof : In this respect, the condition that $\left[v^T \;\middle|\; v_{d,1}^T \;\; \dots \;\; v_{d,N}^T\right] S_b = 0$ is equivalent to saying that $v^T S_{b,x} = 0$. Solving then for $(v_{d,1}^T, \dots, v_{d,N}^T)$ yields the second statement. The proof is complete. ♢

5.3.2 At least one bimolecular reactions enters the queue

We assume here that the first $0 < q < n$ delay input reactions are bimolecular whereas the $n-q$ last input reactions are unimolecular. In this regard, we can partition the stoichiometric matrix associated with bimolecular reactions from queues having a bimolecular reactions as input reactions:

$$S_b = \left[\begin{array}{cc} S_{b,x}^0 & \operatorname*{row}_{i=1}^{q}[S_{b,x}^i] \\ \hline 0 & \operatorname*{diag}_{i=1}^{q}[S_{b,d}^i] \end{array}\right]. \tag{50}$$

Let the matrices B_u and B_b be of the same form as B in (38)-(39) with the restriction that B_u corresponds to queues have unimolecular reactions as input reactions whereas B_b corresponds to queues having bimolecular reactions as input reactions. We use a similar definition for the matrices C_u, H_u and H_b.

We then have the following result:

Theorem 17 *Assume that there exist* $v \in \mathbb{R}_{>0}^d$ *and* $v_{d,i} \in \mathbb{R}^{d_i}$, $i = 1, \dots, q$, *such that the conditions*

$$\left[v^T \;\middle|\; v_{d,1}^T \;\; \dots \;\; v_{d,q}^T\right] S_b = 0 \tag{51}$$

and

$$\left[\begin{array}{c} v \\ \hline v_{d,1} \\ \vdots \\ v_{d,q} \end{array}\right] \left[\begin{array}{c|c} A + B_u H_u^{-T} C_u & B_b \\ \hline 0 & H_b^T \end{array}\right] < 0 \tag{52}$$

hold. Then, the Markov process describing the delayed reaction network $(\boldsymbol{X}, \mathcal{R})$ *is exponentially ergodic and the stationary distribution is light-tailed.*

Proof : The proof follows from the application of Theorem 2. ♢

6 Moments equation

The goal of this section is to describe the influence of phase-type distributed delays on the moment dynamics and their stationary values.

6.1 Dynamics of the first-order moments

The moments equation is given by

$$\frac{d}{dt}\begin{bmatrix}\mathbb{E}[X(t)]\\ \mathbb{E}[D(t)]\end{bmatrix} = \begin{bmatrix}A & B\\ C & H^T\end{bmatrix}\begin{bmatrix}\mathbb{E}[X(t)]\\ \mathbb{E}[D(t)]\end{bmatrix} + \begin{bmatrix}S_x b_x\\ S_{\text{in},d} b_{\text{in}}\end{bmatrix} \tag{53}$$

where the function f and g come from the potential presence of nonlinear propensities. Noting that

$$\mathbb{E}[D(t)] = e^{H^T t}\mathbb{E}[D(0)] + \int_0^t e^{H^T s} C\mathbb{E}[X(t-s)]ds H^{-T}(I - e^{H^T t})S_{\text{in},d} b_{\text{in}} \tag{54}$$

then we obtain the following dynamical model for the moment equations

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= A\mathbb{E}[X(t)] + S_x b_x \\ &\quad + B\left[e^{H^T t}\mathbb{E}[D(0)] + \int_0^t e^{H^T s} C\mathbb{E}[X(t-s)]ds + H^{-T}(I - e^{H^T t})S_{\text{in},d} b_{\text{in}}\right].\end{aligned} \tag{55}$$

We have that

$$\begin{aligned}Be^{H^T s}C &= \int_0^t S_{\text{out},x}\operatorname*{diag}_i(-\mathbb{1}^T H_i^T e^{H_i^T s})\operatorname*{col}_i(\alpha_i^T q_i^T) \\ &= S_{\text{out},x}\operatorname*{col}_i(-\mathbb{1}^T H_i^T e^{H_i^T s}\alpha_i^T q_i^T) \\ &= S_{\text{out},x}\operatorname*{col}_i(-\alpha_i e^{H_i s} H_i \mathbb{1} q_i^T) \\ &= S_{\text{out},x}\operatorname*{col}_i(f_i(s) q_i^T)\end{aligned} \tag{56}$$

where $f_i(s)$ is the probability density function of a random variable with distribution $\text{PH}(H_i, \alpha_i)$. Hence, we get

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= A\mathbb{E}[X(t)] + S_{\text{out},x}\int_0^t \operatorname*{col}_i(f_i(s)q_i^T)\mathbb{E}[X(t-s)]ds + S_x b_x \\ &\quad + B\left[e^{H^T t}\mathbb{E}[D(0)] + H^{-T}(I - e^{H^T t})S_{\text{in},d} b_{\text{in}}\right].\end{aligned} \tag{57}$$

From the above expression, we can see that the first-order moment dynamics does contain convolutions where the first-order moment is convoluted with the delay distributions. In this regard, the presence of phase-type distributed delays in the stochastic dynamics corresponds to a filtering term for the mean dynamics. Interestingly, this connects very well with the fact that if we substitute the convolution kernels by the Dirac distribution, we obtain a standard delay system since

$$\int_0^t \delta(s - \bar{\tau})x(t-s)ds = x(t - \bar{\tau}), t \geq \bar{\tau}. \tag{58}$$

6.2 Invariance of the stationary first-order moments

We have shown that the dynamics of the first-order moments are affected by the presence of phase-type distributed delays. We also know that the delayed network is ergodic provided that its delay-free counterpart is. As a result, the first-order moments converge to a unique equilibrium point. We prove here that the stationary value for the first-order moments coincide with their stationary value in the delay-free network:

Proposition 18 *Assume that the delayed reaction network is ergodic. The stationary first-order moments of the delayed reaction network coincide with the stationary first-order moments of the delay-free reaction network.*

Proof : Assuming ergodicity of the process, then there exists a unique stationary distribution and the limit $\lim_{t\to\infty} \mathbb{E}[X(t)] = \mathbb{E}_\pi[X]$. Furthermore, it solves the

$$\begin{aligned} 0 &= \lim_{t\to\infty} \Bigg(A\mathbb{E}[X(t)] + S_{\text{out},x} \int_0^t \operatorname*{col}_i(f_i(s)q_i^T)\mathbb{E}[X(t-s)]ds + S_x b_x \\ &\quad + B\left[e^{H^T t}\mathbb{E}[D(0)] + H^{-T}(I - e^{H^T t})S_{\text{in},d} b_{\text{in}}\right]\Bigg). \end{aligned} \tag{59}$$

This yields

$$\left(A + S_{\text{out},x} \int_0^\infty \operatorname*{col}_i(f_i(s)q_i^T)ds\right)\mathbb{E}_\pi[X] + S_x b_x + BH^{-T}S_{\text{in},d} b_{\text{in}} = 0 \tag{60}$$

where $BH^{-T} = \operatorname*{diag}_i(\mathbb{1}^T)$. Finally, we get that

$$\left(A + S_{\text{out},x} \operatorname*{col}_i(q_i^T)\right)\mathbb{E}_\pi[X] + S_x b_x + \operatorname*{diag}_i(\mathbb{1}^T)S_{\text{in},d} b_{\text{in}} = 0. \tag{61}$$

The result follows. ♢

This result then demonstrates that phase-type distributed delay do not affect the stationary value of the first-order moments. This is also consistent with the results on time-delay systems.

7 Antithetic integral control of unimolecular reaction network with phase-type distributed delays

We now address the problem of controlling a delayed stochastic reaction network using an antithetic integral controller. Interestingly, it can be shown that delays can also affect the reactions of the controller without requiring the development of a new theory. We now consider the following delayed antithetic integral controller

$$\boldsymbol{Z_1} \xrightarrow{k,\tau(H_a,\alpha_a)} \boldsymbol{Z_1} + \boldsymbol{X_1}, \boldsymbol{X_\ell} \xrightarrow{\theta,\tau(H_m,\alpha_m)} \boldsymbol{X_\ell} + \boldsymbol{Z_2}, \emptyset \xrightarrow{\mu,\tau(H_m,\alpha_m)} \boldsymbol{Z_1}, \boldsymbol{Z_2} + \boldsymbol{Z_1} \xrightarrow{\eta} \emptyset. \tag{62}$$

It is important to stress that the annihilation reaction is not delayed since it is a degradation reaction. Adding a delay line here will not change the behavior of the system as the compound will simply follow the delay line and be degraded in the end. However, from the controller point of view, degradation already occurred when the compound entered the delay line.

We assume that the actuated species $\boldsymbol{X_1}$ are produced after the stochastic time-varying delay modeled by the first delay-line. Hence, we have that $q_1 = ke_1$. Also, we have that the first column of $S_{\text{out},x}$ is e_1, as we produce $\boldsymbol{X_1}$. Similarly, the sensing species is produced after the stochastic time-varying delay modeled by the last delay-line; i.e. the N-th delay line. Hence, we have that $q_N = e_\ell$. We have the following result:

Theorem 19 *Assume that the state-space of the delayed reaction network $(\boldsymbol{X}, \mathcal{R})$ is irreducible and that the network only contains first-order reactions with mass-action kinetics. Then, the following statements are equivalent:*

(a) There exists some vectors $v \in \mathbb{R}^{d+d_1+\ldots+d_N}_{>0}$ and $w \in \mathbb{R}^{d+d_1+\ldots+d_N}_{\geq 0}$ such that the conditions

$$v^T \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] < 0, \tag{63}$$

$$w^T \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] + \left[\begin{array}{c|ccc} 0 & 0 & \ldots & -\mathbb{1}^T H_N^T \end{array}\right] = 0 \tag{64}$$

and

$$\left[\begin{array}{c|ccc} 0 & \alpha_1 & 0 & \ldots & 0 \end{array}\right] w > 0 \tag{65}$$

hold.

(b) There exists some vectors $v \in \mathbb{R}^d_{>0}$ and $w \in \mathbb{R}^d_{\geq 0}$, $w_1, w_\ell > 0$, such that the conditions

$$v^T(A - BH^{-T}C) < 0 \text{ and } w^T(A - BH^{-T}C) + e_\ell^T = 0 \tag{66}$$

hold.

(c) The controlled stochastic reaction network with delays is ergodic and we have that $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$.

Proof : The conditions in the first statement are nothing else but the conditions for the delayed reaction network. To prove the equivalence between the conditions we just do some algebra. First, note that we have that

$$\bar{v}^T \begin{bmatrix} A & B \\ C & H^T \end{bmatrix} < 0 \tag{67}$$

if and only if $v^T(A - BH^{-T}C) < 0$. We also have that

$$\bar{w}^T \begin{bmatrix} A & B \\ C & H^T \end{bmatrix} + \left[\begin{array}{c|ccc} 0 & 0 & \ldots & -\mathbb{1}^T H_N^T \end{array}\right] = 0 \tag{68}$$

together with

$$\bar{w}^T \left[\begin{array}{c} 0 \\ \hline \alpha_1^T k \mathbb{E}_\pi[Z_1] \\ 0 \\ \vdots \\ 0 \end{array}\right] > 0. \tag{69}$$

Let $\bar{\bar{w}} = (\bar{\bar{w}}_1, \bar{\bar{w}}_2)$, where $\bar{w}_1 \in \mathbb{R}^d_{>0}$ and $\bar{w}_2 \in \mathbb{R}^{d_1+\ldots+d_N}_{\geq 0}$. Hence, from $\bar{\bar{w}}_1^T B + \bar{w}_2^T H^T = 0$, we get that

$$\bar{w}_2^T \;=\; -(\left[0 \;\middle|\; 0 \;\; \ldots \;\; -\mathbb{1}^T H_N^T\right] + \bar{w}_1^T B) H^{-T} \tag{70}$$

and hence

$$\bar{w}_1^T (A - BH^{-T}C) + \left[0 \;\; \ldots \;\; -\mathbb{1}^T H_N^T\right] H^{-T} C = 0 \tag{71}$$

which is equivalent to the condition that $\bar{w}_1^T (A - BH^{-T}C) + e_\ell^T = 0$. Using this value for $\bar{w}_2$, we also get that

$$\begin{aligned} \bar{w}^T \left[\begin{array}{c} 0 \\ \hline \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{array}\right] &= \bar{w}_2^T \begin{bmatrix} \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} = -(\left[0 \;\middle|\; 0 \;\; \ldots \;\; -\mathbb{1}^T H_N^T\right] + \bar{w}_1^T B) H^{-T} \begin{bmatrix} \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= -\bar{w}_1^T S_{\text{out},x} \operatorname*{diag}_i(\mathbb{1}^T) H^{-T} \begin{bmatrix} \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= -\bar{w}_1^T e_1 \mathbb{1}^T H_1^{-T} \alpha_1^T \\ &= -\bar{w}_1^T e_1 \end{aligned} \tag{72}$$

where we have used the fact that $\mathbb{1}^T H_1^{-T} \alpha_1^T = 1$. We have proven the equivalence between the statements. However, it remains to prove that $\mathbb{E}[X_\ell(t) \to \mu/\theta$. To do so, we first write down the moments equation

$$\begin{aligned} \frac{d}{dt} \left[\begin{array}{c} \mathbb{E}[X(t)] \\ \hline \mathbb{E}[D(t)] \end{array}\right] &= \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] \begin{bmatrix} \mathbb{E}[X(t)] \\ \mathbb{E}[D(t)] \end{bmatrix} + \left[\begin{array}{c} 0 \\ \hline k\alpha_1^T \\ 0 \\ \vdots \\ 0 \end{array}\right] \mathbb{E}[Z_1(t)] \\ \frac{d}{dt}\mathbb{E}[Z_1(t)] &= \mu - \eta \mathbb{E}[Z_1(t)]\mathbb{E}[Z_2(t)] \\ \frac{d}{dt}\mathbb{E}[Z_2(t)] &= \left[0 \;\middle|\; 0 \;\; \ldots \;\; 0 \;\; -\theta \mathbb{1}^T H_N^T\right] \begin{bmatrix} \mathbb{E}[X(t)] \\ \mathbb{E}[D(t)] \end{bmatrix} - \eta \mathbb{E}[Z_1(t)]\mathbb{E}[Z_2(t)]. \end{aligned} \tag{73}$$

The stationary solution is given by

$$\begin{bmatrix}\mathbb{E}_\pi[X]\\ \mathbb{E}_\pi[D]\end{bmatrix} = -k\begin{bmatrix}A & B\\ C & H^T\end{bmatrix}^{-1}\begin{bmatrix}0\\ \hline \alpha_1^T k\mathbb{E}_\pi[Z_1]\\ 0\\ \vdots\\ 0\end{bmatrix} \tag{74}$$

and

$$\mu/\theta = \left[0 \;\middle|\; 0 \;\; \dots \;\; -\mathbb{1}^T H_N^T\right]\begin{bmatrix}\mathbb{E}_\pi[X]\\ \mathbb{E}_\pi[D]\end{bmatrix} \tag{75}$$

where $\mathbb{E}_\pi$ denotes the expectation operator at stationarity. We use now the following formula for the inversion of block-matrices

$$\begin{bmatrix}A & B\\ C & H^T\end{bmatrix}^{-1} = \begin{bmatrix}(A-BH^{-T}C)^{-1} & (A-BH^{-T}C)^{-1}BH^{-T}\\ -H^{-T}C(A-BH^{-T}C)^{-1} & H^{-T}+H^{-T}C(A-BH^{-T}C)^{-1}BH^{-T}\end{bmatrix} \tag{76}$$

where

$$\begin{aligned} A-BH^{-T}C &= A+\sum_i S^i_{\text{out},x}q_i^T\\ H^{-T}C &= \begin{bmatrix}H_1^{-T}\alpha_1^T q_1^T\\ \vdots\\ H_N^{-T}\alpha_N^T q_N^T\end{bmatrix}\\ BH^{-T} &= = S_{\text{out},x}\operatorname*{diag}_i(\mathbb{1}^T)\end{aligned} \tag{77}$$

Finally, we have that

$$-\left[0 \;\middle|\; 0 \;\; \dots \;\; -\mathbb{1}^T H_N^T\right]\begin{bmatrix}A & B\\ C & H^T\end{bmatrix}^{-1}\begin{bmatrix}0\\ \hline \alpha_1^T k\mathbb{E}_\pi[Z_1]\\ 0\\ \vdots\\ 0\end{bmatrix} = \mu/\theta \tag{78}$$

and, hence, $\mathbb{E}_\pi[Z_1] = \dfrac{\mu}{g\theta}$ where $g := e_\ell^T(A-BH^{-T}C)^{-1}e_1$, where we have used the fact that $S_{\text{out},x}e_1 = e_1$. Note, moreover, that g is the static-gain of the non-delayed network. Since we have that $\mathbb{E}_\pi[X_\ell] = gk\mathbb{E}_\pi[Z_1]$, then we obtain that

$$\mathbb{E}_\pi[X_\ell] = \mu/\theta. \tag{79}$$

This proves that $\boldsymbol{X_\ell}$ is also regulated and completes the proof. $\diamondsuit$

As for the ergodicity analysis, we can see that the conditions also reduce to the condition for delay-free networks.

8 Concluding statements

Delays that are phase-type distributed have been shown to be natural to consider in the context of stochastic reaction networks. In fact, stochastic time-varying delays can be represented as reaction networks with conversion reactions. In this regard, any reaction network with delayed reactions having a delay that is phase-type distributed can be equivalently represented by a reaction network with augmented state-space. Yet, the dimension of the state-space remains finite. In this regard, existing results for the analysis and the control of stochastic reaction networks remain applicable and do not need to be extended to the infinite-dimensional case – a difficult task. We first characterize all the delays that are phase-type distributed in terms of algebraic conditions. We then provide an explicit way for building the associated reaction network and interconnect it to the original network. Ergodicity tests are then provided and it proves that for unimolecular networks, the delayed reaction network is ergodic if and only if the delay-free network is ergodic as well. In this regard, delays are not harmful to the ergodicity property. This also indicate that delays can be arbitrarily complicated as long as they are phase-type distributed. For bimolecular networks, the situation is more complex but it can be shown that in certain cases the ergodicity conditions of the delayed network are fulfilled if and only if the conditions are fulfilled for the delay-free network. The analysis of the first moment equation demonstrate that the delays yield additional convolution terms in the mean dynamics of the molecular species. However, they do not change the stationary mean values and the stationary means are the same as in the delay-free case. Finally, the antithetic integral control of such networks is addressed. The controller is also extended to incorporate delays. It is shown that the ergodicity and output controllability conditions on the delayed reaction network are satisfied if and only if they are satisfied in the delay-free case.

All the obtained results can be straightforwardly extended to deal with uncertain reaction rates as in [28–30]. This was not done here because this extension is immediate and would not bring much to this paper. Note that it is unnecessary to consider uncertain reaction rates for the reactions in the delay lines as the results are independent of these parameters. Robustness with all delays is indeed automatically ensured whenever the (anyway necessary) condition that the delay-free network be ergodic is satisfied.

Extension to delay lines with finite capacity would be a very interesting topic to consider in order to consider the use of finite resources like in enzymatic networks [57] or some queueing processes that mRNA undergo while leaving the nucleus [58]. This can be done by incorporating counter species counting the number of elements in the queue and inactivating the input reaction whenever the queue is full. The difficulty here resides in the fact that some of the queuing reactions become nonlinear and, hence, more difficult to consider. However, approximations could be helpful here in obtaining interesting results.

The proposed approach allows one to consider almost deterministic constant delays in the sense that one can consider phase-type distributions that are arbitrarily close to the Dirac distribution without destroying the ergodicity of the delayed reaction network provided that its delay-free counterpart is ergodic. However, if constant deterministic delays need to be exactly incorporated, a new theory may be needed, which is definitely of interest. A starting

point would be the consideration of a Markov jump processes on an infinite-dimensional state-space. The difficulty here will be the verification of the irreducibility property of the, now infinite-dimensional, state-space. A potential idea would be to look at reachability results for time-delay systems or, more generally, for infinite-dimensional systems described by partial differential equations. Regarding the verification of the positive recurrence property of the Markov process, the use of Foster-Lyapunov functions may not be suitable. Instead, we might need to consider more general functions or functionals. A possible starting point would be to look, again, at the literature on time-delay systems and adapt important tools such as Lyapunov-Razumikhin functions or Lyapunov-Krasovskii functionals to out present context.

References

[1] V. B. Kolmanovskii and A. D. Myshkis, *Applied Theory of functional differential equations.* Kluwer, 1992.

[2] S. I. Niculescu, *Delay effects on stability. A robust control approach.* Springer-Verlag: Heidelbeg, 2001, vol. 269.

[3] K. Gu, V. L. Kharitonov, and J. Chen, *Stability of Time-Delay Systems.* Birkhäuser, Boston, 2003.

[4] W. Michiels and S. I. Niculescu, *Stability and stabilization of time-delay systems. An eigenvalue based approach.* Philadelphia, USA: SIAM Publication, 2007.

[5] C. Briat, *Linear Parameter-Varying and Time-Delay Systems – Analysis, Observation, Filtering & Control*, ser. Advances on Delays and Dynamics. Heidelberg, Germany: Springer-Verlag, 2015, vol. 3.

[6] E. Fridman, *Introduction to Time-Delay Systems.* Springer International Publishing Basel Switzerland: Birkhäuser, 2014.

[7] K. Gopalsamy, *Stability and oscillations in delay differential equations of population dynamics*, ser. Mathematical and Its Applications. Research Triangle Park, North Carolina, USA: Kluwer Academic Publishers, 1992, vol. 74.

[8] K. Gopalsamy and B. G. Zhang, "On a neutral delay-logistic equation," *Dynamics and stability of systems: An international journal*, vol. 2(3-4), pp. 183–195, 1988.

[9] C. Briat and E. I. Verriest, "A new delay-SIR model for pulse vaccination," *Biomedical signal processing and control*, vol. 4(4), pp. 272–277, 2009.

[10] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, "Delay-induced stochastic regulation oscillations in gene regulation," *Proceedings of the National Academy of Sciences*, vol. 102(42), pp. 14 593–14 598, 2005.

[11] M. Jansen and P. Pfaffelhuber, "Stochastic gene expression with delays," *Journal of Theoretical Biology*, vol. 364, pp. 355–363, 2015.

[12] E. Fridman, C. Bonnet, F. Mazenc, and W. Djema, "Stability of the cell dynamics in acute myeloid leukemia," *Systems & Control Letters*, vol. 88, pp. 91–100, 2016.

[13] W. Djema, F. Mazenc, and C. Bonnet, "Stability analysis and robustness results for a nonlinear system with distributed delays describing hematopoiesis," *Systems & Control Letters*, vol. 102, pp. 93–101, 2017.

[14] W. Djema, C. Bonnet, F. Mazenc, J. Clairambault, E. Fridman, P. Hirsch, and F. Delhommeau, "Control in dormancy or eradication of cancer stem cells: Mathematical modeling and stability issues," *Journal of Theoretical Biology*, vol. 449, pp. 103–123, 2018.

[15] V. B. Kolmanovskii and V. R. Nosov, *Stability of functional differential equations.* Academic Press, London, 1986.

[16] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," *Nature*, vol. 456, pp. 516–519, 2008.

[17] W. H. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, "Delay-induced degrade-and-fire oscillations in small genetic circuits," *Physical Review Letter*, vol. 102, p. 068105, 2009.

[18] K. Josić, J. M. López, W. Ott, L. Shiau, and M. R. Bennett, "Stochastic delay accelerates signaling in gene networks," *PLOS Computational Biology*, vol. 7(11), p. e1002264, 2011.

[19] M. Scott, T. Hawa, and B. Ingalls, "Deterministic characterization of stochastic genetic circuits," *Proc. Natl. Acad. Sci.*, vol. 104(18), pp. 7402–7407, 2007.

[20] D. Anderson and T. G. Kurtz, "Continuous time Markov chain models for chemical reaction networks," in *Design and analysis of biomolecular circuits - Engineering Approaches to Systems and Synthetic Biology*, H. Koeppl, D. Densmore, G. Setti, and M. di Bernardo, Eds. Springer Science+Business Media, 2011, pp. 3–42.

[21] D. F. Anderson, "A modified next reaction method for simulating chemical systems with time-dependent propensities and delays," *Journal of Chemical Physics*, vol. 127, p. 214107, 2007.

[22] C. C. Guet, A. Gupta, A. Henzinger, M. Mateescu, and A. Sezgin, "Delayed continuous-time markov chains for genetic regulatory circuits," in *Computer Aided Verification*, ser. Lecture Notes in Computer Science, P. Madhusudan and S. A. Seshia, Eds. Springer Berlin Heidelberg, 2012, vol. 7358, pp. 294–309.

[23] M. Barrio, A. Leier, and T. T. Marquez-Lago, "Reduction of chemical reaction networks through delay distributions," *The Journal of Chemical Physics*, vol. 138, p. 104114, 2013.

[24] A. Leier, M. Barrio, and T. T. Marquez-Lago, "Exact model reduction with delays: closed-form distributions and extensions to fully bi-directional monomolecular reactions," *Interface*, vol. 11, p. 20140108, 2014.

[25] J. Falk, L. Bronstein, M. Hanst, B. Drossel, and H. Koeppl, "Context in synthetic biology: Memory effects of environments with mono-molecular reactions," *ArXiv e-prints*, 2018.

[26] A. Gupta, C. Briat, and M. Khammash, "A scalable computational framework for establishing long-term behavior of stochastic reaction networks," *PLOS Computational Biology*, vol. 10(6), p. e1003669, 2014.

[27] C. Briat, A. Gupta, I. Shames, and M. Khammash, "Scalable tests for ergodicity analysis of large-scale interconnected stochastic reaction networks," in *21st International Symposium on Mathematical Theory of Networks and Systems*, Groningen, The Netherlands, 2014, pp. 92–95.

[28] C. Briat and M. Khammash, "Robust ergodicity and tracking in antithetic integral control of stochastic biochemical reaction networks," in *55th IEEE Conference on Decision and Control*, Las Vegas, USA, 2016, pp. 752–757.

[29] ——, "Robust and structural ergodicity analysis of stochastic biomolecular networks involving synthetic antithetic integral controllers," in *20th IFAC World Congress*, Toulouse, France, 2017, pp. 11 405–11 410.

[30] C. Briat, "Sign properties of Metzler matrices with applications," *Linear Algebra and its Applications*, vol. 515, pp. 53–86, 2017.

[31] A. Gupta and M. Khammash, "Computational identication of irreducible state-spaces for stochastic reaction networks," *SIAM Journal of Applied Dynamical Systems*, vol. 17(2), pp. 1213–1266, 2018.

[32] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems. Queueing theory in action.* New York, USA: Cambridge University Press, 2013.

[33] M. Bladt, "A review of phase-type distributions and their use in risk theory," *ASTIN Bulletin*, vol. 35(1), pp. 145–161, 2005.

[34] B. Shaw and A. H. Marshall, *Modelling the total time spent in an Accident and Emergency department and the associated costs.* World Scientific, 2007, ch. 21, pp. 172–180.

[35] K. Strimmer and O. G. Pybus, "Exploring the demographic history of dna sequences using the generalized skyline plot," *Molecular Biology and Evolution*, vol. 18(12), pp. 2298–2305, 2001.

[36] S. Asmussen, *Applied Probability and Queues*. New York, USA: Springer-Verlag, 2003.

[37] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distributions via the EM algorithm," *Scandinavian Journal Journal of Statistics*, vol. 23(4), pp. 419–441, 1996.

[38] C. Briat, A. Gupta, and M. Khammash, "Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks," *Cell Systems*, vol. 2, pp. 17–28, 2016.

[39] T. Jahnke and W. Huisinga, "Solving the chemical master equation for monomolecular reaction systems analytically," *Journal of Mathematical Biology*, vol. 54, pp. 1–26, 2007.

[40] V. Kazeev, M. Khammash, M. Nip, and C. Schwab, "Direct solution of the chemical master equation using quantized tensor trains," *PLOS Computational Biology*, vol. 10(3), p. e1003359, 2014.

[41] A. Gupta, J. Mikelson, and M. Khammash, "A finite state projection algorithm for the stationary solution of the chemical master equation," *Journal of Chemical Physics*, vol. 147, no. 154101, pp. 1–23, 2017.

[42] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, pp. 287–289, 1977.

[43] R. D. Bliss, P. R. Painter, and A. G. Marr, "Role of feedback inhibition in stabilizing the classical operon," *Journal of Theoretical Biology*, vol. 97, pp. 177–193, 1982.

[44] L. Glass, A. Beuter, and D. Larocque, "Time delays, oscillations, and chaos in physiological control systems," *Mathematical Biosciences*, vol. 90, pp. 111–125, 1988.

[45] S. Ruan., "Delay differential equations in single species dynamics," in *Delay Differential Equations and Applications*, O. Arino, M. L. Hbid, and E. Ait Dads, Eds. Springer Netherlands, 2006, pp. 477–517.

[46] S. P. Meyn and R. L. Tweedie, "Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes," *Adv. Appl. Prob.* , vol. 25, pp. 518–548, 1993.

[47] A. Milias-Argeitis and M. Khammash, "Optimization-based lyapunov function construction for continuous-time markov chains with afne transition rates," in *53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA, 2014, pp. 4617–4622.

[48] S. Engblom, "On the stability of stochastic jump kinetics," *ArXiv:1202.3892*, 2012. [Online]. Available: http://arxiv.org/abs/1202.3892

[49] M. Rathinam, "Moment growth bounds on continuous time Markov processes on non-negative integer lattices," *Quaterly of Applied Mathematics*, vol. LXXIII, no. 2, pp. 347–364, 2015.

[50] D. F. Anderson and J. Kim, "Some network conditions for positive recurrence of stochastically modeled reaction networks," *SIAM Journal on Applied Mathematics*, vol. 78(5), pp. 2692–2713, 2018.

[51] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, MA, USA: Cambridge University Press, 2004.

[52] G. Lillacci, S. K. Aoki, D. Schweingruber, and M. Khammash, "A synthetic integral feedback controller for robust tunable regulation in bacteria," *bioRxiv*, 2017. [Online]. Available: http://www.biorxiv.org/content/early/2017/08/01/170951

[53] J. E. Marshall, *Control of Time-Delay Systems*. Peter Peregrinus, Stevenage, 1979.

[54] P. J. Gawthrop and M. T. Nihtila, "Identification of time delays using a polynomial identification method," *Systems & Control Letters*, vol. 5, pp. 267–271, 1985.

[55] G. H. Tomlinson and M. J. Somerville, "Low-pass approximations to a time-delay characteristic," *International Journal of Electronics*, vol. 15(5), pp. 495–512, 1963.

[56] J. Lam, "Convergence of a class of padé approximations for time-delay systems," *International Journal of Control*, vol. 52(4), pp. 989–1008, 1990.

[57] P. J. Steiner, R. J. Williams, J. Hasty, and L. S. Stimring, "Criticality and adaptivity in enzymatic networks," *Biophysical Journal*, vol. 111, pp. 1078–1087, 2016.

[58] M. M. K. Hansen, R. V. Desai, M. L. Simpson, and L. S. Weinberger, "Cytoplasmic amplication of transcriptional noise generates substantial cell-to-cell variability," *Cell Systems*, vol. 7, pp. 384–397, 2018.