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

Corentin Briat*and Mustafa Khammash†

Abstract

Delays are important phenomena arising in a wide variety of real world systems, including biological ones, because of diffusion/propagation effects or as simplifying modeling elements. We propose here to consider delayed stochastic reaction networks, a class of networks that has been relatively few studied until now. The difficulty in analyzing them resides in the fact that their state-space is infinite-dimensional. We demonstrate here that by restricting the delays to be phase-type distributed, one can represent the associated delayed reaction network as a reaction network with finite-dimensional state-space. This can be achieved by suitably adding chemical species and reactions to the delay-free network following a simple algorithm which is fully characterized. Since phase-type distributions are dense in the set of probability distributions, they can approximate any distribution arbitrarily closely and this makes their consideration only a bit restrictive. As the state-space remains finite-dimensional, usual tools developed for non-delayed reaction network directly apply. In particular, we prove, for unimolecular mass-action reaction networks, that the delayed stochastic reaction network is ergodic if and only if the delay-free network is ergodic as well. Bimolecular reactions are more difficult to consider but slightly stronger analogous results are nevertheless obtained. These results demonstrate that delays have little to no harm to the ergodicity property of reaction networks as long as the delays are phase-type distributed, and this holds regardless the complexity of their distribution. 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. The covariance, however, is influenced by the presence of the delays. 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.

*D-BSSE, ETH-Zürich, Switzerland (*corentin@briat.info*, www.briat.info).

†D-BSSE, ETH-Zürich, Switzerland (*mustafa.khammash@bsse.ethz.ch*, https://www.bsse.ethz.ch/ctsb.)

1 Introduction

Delays are omnipresent physical phenomena induced by memory, propagation or transport effects [12, 24, 30, 42, 50, 52]. They naturally arise in population dynamics [28], ecology [29], epidemiology [19], biology [11, 20, 21, 25, 39] and engineering [12, 30, 43, 52]. 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, for instance, to design oscillators [11, 48, 58]. In the stochastic setting, delays have indeed been shown to be helpful for generating oscillations [11], but also to accelerate signaling [40] and to be responsible for an increase in intrinsic variability [55]. 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. [12, 30, 42, 50, 52]. When the dynamics of the reaction network is inherently stochastic and represented by a continuous-time jump Markov process [1], 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 [11], the next reaction method [2] or using delayed continuous-time Markov chains [31]. Alternatively, it has been shown in [7, 23, 45] that certain chains of unimolecular reactions in a reaction network could be equivalently substituted by stochastic time-varying delays whose distributions can exactly computed from the reaction rates. This has led to a drastic speedup in the stochastic simulations.

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 [13, 16–18, 32, 33]. Ergodicity is the stochastic analogue of having a unique globally attractive fixed point for deterministic dynamics. It can 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 to establishing two properties: the irreducibility of the state-space (or a particular subset of it) and the fulfilment of a Foster-Lyapunov condition. While it is quite clear how these conditions could be checked for standard (i.e. non-delayed) 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 not easy to generalize but tools from time-delay systems theory, such as Lyapunov-Krasovskii functionals, may need turn out to be useful in this task.

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 the consideration of a Markov jump system with infinite-dimensional state-space and remain in the finite-dimensional case. 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 [37], risk theory [8], health-care [56] and evolution [59]. As those distributions are dense (in the sense of weak convergence) in the set of all probability distributions on $(0, \infty)$ [5], they can be used to approximate any delay-distribution with arbitrary precision. In this regard, this class of distributions are theoretically mildly restrictive if we allow those distributions to have an arbitrarily large number of free parameters. Several algorithms are available for approximating a given distribution or for fitting empirical ones; see e.g. [6]. The reason behind the use of such distributions is that they can be exactly represented as an irreducible unimolecular reaction network, meaning that delays can be included in the network by suitably adding extra species and extra reactions, thereby creating an augmented network with a finite, yet possibility high-, dimensional state-space. A procedure for characterizing distributions that can be represented as phase-type ones is proposed along with a constructive method for representing the phase-type distributed delay as a minimal reaction network. Such a minimal network is notably shown to be unique. As a consequence, since the state-space remains finite-dimensional, existing tools can be applied to the augmented network to yield results on delayed reaction networks having delays that are phase-type distributed.

We propose to use the tools developed in [15, 32] in order to establish several results for delayed reaction networks. Using the ergodicity results developed in [32], we prove that a delayed unimolecular network 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 the complexity of their distribution. This includes complex distributions approximating arbitrarily closely heavy-tailed distributions or Dirac distributions. In this regard, the computational complexity of checking whether a delayed reaction network is ergodic does not increase as the delay-distribution increases in complexity but only depends on the number of nominal molecular species involved in the delayed network. This result parallels existing ones on the stability of linear positive systems with delays for which it is well-known that the stability is equivalent to that of the delay-free system and is independent of the value of the delay; see e.g. [14, 35].

Similar results are obtained for bimolecular networks. When there is no delayed bimolecular reactions, it is shown that the ergodicity conditions reduces to those of the delay-free network. So, in this case again, one can see that the delays are not detrimental to the ergodicity of the network and that the complexity of verifying whether a delayed reaction network satisfying those conditions is the same as checking the ergodicity of a bimolecular network. When some delayed bimolecular reactions are present, the conditions of the delay-free case are not fully recovered but only a slight variations of them, not necessarily more restrictive. In this case also, the results can be associated with an ergodicity test for a delay-free network having a lower computational complexity than the augmented one.

It is then shown that phase-type distributed delays introduce convolution terms in the dynamics of the species of the delayed reaction network with kernels corresponding to the distributions of the delays. Interestingly, we also prove, under the ergodicity assumption,

that the mean stationary value coincides with that of the delay-free network.

Finally, we address the problem of controlling delayed reaction networks using the antithetic integral controller proposed in [15]. We notably generalize the controller to include delayed reactions in the actuation and the measurement parts of the controller. 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 if all its eigenvalues have negative real part. The matrix M is said to be Metzler if all its off-diagonal elements are nonnegative. The standard basis of $\mathbb{R}^n$ is denoted by the vectors $e_1, \ldots, e_n$.

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 through 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 [32].

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 $p(0,x) = \delta(x-x_0)$ is the Kronecker δ function. In general, the CME is not analytically solvable except in some particular simple cases; see e.g. [38] and the references therein. This is the reason why numerical solutions are of interest; see e.g. [34, 41]. 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 [9,27,28,46,54] and stochastic [2,11,39] 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 $\zeta_k^r \in \mathbb{Z}^d_{\geq 0}$ and $\zeta_k^\ell \in \mathbb{Z}^d_{\geq 0}$ are the right- and left-stoichiometric vector of the k-th reaction. Then, we can decompose each delayed reaction 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 [1]. 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. However, depending on the considered network, the above decomposition may not be the best one to represent a delayed reaction as the species on the left-hand side on the reaction are not necessarily destroyed (e.g. in catalytic reactions). This will be further discussed in Section 4.

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 purposes and incorporating delays for modeling and analysis purposes. The first one is usually easier than the second one as the problem is of computational nature and existing algorithms can 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, makes irreducibility analysis of the network more difficult.

Deterministic and stochastic constant delays. Certainly the most natural type of delays that comes to mind are deterministic constant delays. As briefly stated in [1], 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. Additionally, the ergodicity of such deterministically time-varying system may not be easy to define.

Stochastic time-varying delays. Stochastic time-varying delays, as we shall see later, are much more natural to consider in this context provided that their distribution is of some particular kind. It has been shown in [7, 45] 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. [5]. It will be notably demonstrated that delayed reaction networks can be exactly reformulated as a non-delayed network with finite-dimensional state-space.

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. This is formalized in the definition below:

Definition 1 *Let us consider the stochastic reaction network* $(\boldsymbol{X}, \mathcal{R})$*. The associated Markov process, described by the CME* (2)*, is ergodic if there exists a unique stationary distribution* p^* *such that*

$$p_{x_0}(t, x) \to p^*(x)$$

as $t \to \infty$ *for all initial conditions* $p(0, x) = \delta(x - x_0)$*,* $x_0 \in \mathbb{Z}^d_{\geq 0}$*. Moreover, when the convergence is exponentially fast, then the Markov process is* exponentially ergodic*.*

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 [49]. This approach has been specialized to the case of stochastic reaction networks in [32, 51]. Additional results on the stability of reaction networks have also been provided in [3,22,53]. We have the following general result that we specialize to our setup [49]:

Theorem 2 *Let us consider the stochastic reaction network* $(\boldsymbol{X}, \mathcal{R})$ *and assume that its state-space is irreducible. Let* $\mathbb{A}$ *be the generator of the underlying Markov process as defined in* (4) *and let* $C \subset \mathbb{Z}^d_{\geq 0}$ *be some compact set. Assume further that there exists a function* $V : \mathbb{Z}^d_{\geq 0} \mapsto \mathbb{R}$*, a function* $f : \mathbb{Z}^d_{\geq 0} \mapsto \mathbb{R}$*,* $f(x) \geq 1$ *for all* $x \in \mathbb{Z}^d_{\geq 0}$ *and scalars* $c, d > 0$ *such that*

(a) V *is norm-like (i.e. nonnegative and radially unbounded)*

(b) $\mathbb{A}V(x) \leq -cf(x) + d1_C(x)$ *where* 1_C *is the indicator function of the set* C*.*

Then, the Markov process is ergodic where the convergence is in total variation. Moreover, if the above conditions hold with $f(x) = V(x)$*, then the Markov process is exponentially ergodic. In any of those cases, the function* V *is said to be Foster-Lyapunov function.*

Equivalent conditions for the exponential ergodicity are $V(0) = 0$, $V(x) > 0$ for all $x \in \mathbb{R}^d$, V radially unbounded and $\mathbb{A}V(x) \leq c_1 - c_2 V(x)$ for some $c_1 \geq 0, c_2 > 0$ and for all $x \in \mathbb{Z}^d_{\geq 0}$. This reformulation similar to standard Lyapunov conditions will be preferred in this paper for simplicity.

In the case of stochastic mass-action reaction networks with zeroth-, first- and second-order reactions, it is possible to exploit the positivity of the dynamics of the system and consider a linear copositive 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

$$S = \begin{bmatrix} S_0 & S_u & S_b \end{bmatrix} \in \mathbb{Z}^{d \times K} \tag{7}$$

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 $\lambda_0 \geq 0$ and $\lambda_u(x) = W_u x$ where W_u is an appropriate nonnegative real-valued matrix. We also make the following assumption:

Assumption 3 *The network is open in the sense that it has no closed component and the state-space of the associated Markov process is irreducible.*

The assumption on the openness of the network is considered to simplify the exposition of the results but does not make them more restrictive since establishing the ergodicity of closed-components simply follows from the irreducibility of their state-space. The irreducibility is also assumed here as this is not the main topic of this paper. However, it seems then important to clarify how this 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 positive values of the reaction rates. Zero values, if considered, will need to be taken care separately in order to make sure that the removal of each of the corresponding reactions do not destroy the irreducibility property of the state-space. It has been proven that the irreducibility of reaction networks can be checked by solving a linear program and a simple linear algebraic condition [33]. 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 [32] that they have successfully been able to establish the ergodicity of several typical reaction networks considered in the literature.

We first consider the case unimolecular networks and provide the following result which consists of a slight generalization of the one in [32]:

Theorem 4 (Unimolecular reaction networks) *Let us consider stochastic reaction network $(\boldsymbol{X}, \mathcal{R})$ with zeroth- and first-order mass-action kinetics which satisfies Assumption 3. Then, the following statements are equivalent:*

(a) There exists a vector $v \in \mathbb{R}^d_{>0}$ for which the function $V(x) = v^T x$ is a Foster-Lyapunov function for the corresponding Markov process.

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

(c) *The matrix $S_u W_u$ is Hurwitz stable (i.e. all the eigenvalues have negative real-part).*

(d) *The Markov process describing the reaction network is exponentially ergodic.*

Moreover, the stationary distribution is light-tailed and, as a consequence, all the moments are bounded and globally converging to their unique stationary value.

Proof : To prove the equivalence between the statements (a) and (b), we consider the modified versions of the conditions in Theorem 2 with the function $V(x) = v^T x$, $v > 0$. We get that

$$\mathbb{A}V(x) = v^T(S_u W_u x + S_0 \lambda_0). \tag{8}$$

Clearly if $v^T S_u W_u < 0$, then there exists a $c_2 > 0$ such that $\mathbb{A}V(x) \leq -cV(x) + c_1$ where $c_1 = v^T S_0 \lambda_0$. This proves the equivalence. The equivalence with the statement (c) comes from the fact that $S_u W_u$ is Metzler and that the condition in the statement (b) is a necessary and sufficient condition for its Hurwitz stability. The implication of the statement (d) from the statement (a) follows from Theorem 2. To prove the implication of the statement (c) from the statement (d), first note that the dynamics of the first-order moments is given by

$$\frac{d\mathbb{E}[X(t)]}{dt} = S_u W_u \mathbb{E}[X(t)] + S_0 \lambda_0. \tag{9}$$

Since the network is (exponentially) ergodic, then the first order moments need to converge to a unique steady-state value and, as the network has no closed components, this can only be the case if the matrix $S_u W_u$ is Hurwitz stable. The conclusion on the light-tailedness follows from [32]. ♢

Remark 5 *When the network has irreducible closed components, one can simply remove those components from the network and apply the above result to the reduced network. In this case, the matrix $S_u W_u$ will correspond to the matrix describing the interactions between the species in the open part of the network.*

The case of bimolecular networks is a bit more involved but one existing result shows that their analysis may not be too far off from the analysis of unimolecular networks [32]:

Theorem 6 (Bimolecular reaction networks) *Let us consider a stochastic reaction network $(\boldsymbol{X}, \mathcal{R})$ with zeroth-, first- and second-order mass-action kinetics satisfying Assumption 3. Assume further that one of the following equivalent statements holds:*

(a) *There exists a $v \in \mathbb{R}^d_{>0}$ such that $v^T S_b = 0$ for which the function $V(x) = v^T x$ is a Foster-Lyapunov function for the corresponding Markov process.*

(b) *There exists a $v \in \mathbb{R}^d_{>0}$ such that the conditions*

$$v^T S_u W_u < 0 \text{ and } v^T S_b = 0 \tag{10}$$

hold.

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

Remark 7 *If the condition $v^T S_b = 0$ is relaxed to $v^T S_b \leq 0$ where at least one entry is nonzero, we cannot conclude anymore on the light-tailedness of the stationary distribution but the exponentially ergodicity property still holds.*

Theorem 4 and Theorem 6 are interesting for different reasons. The first one is that they can be stated in a very accessible way using elementary linear algebra concepts. 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 efficient algorithms exist; see e.g. [10]. 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.

2.5 Antithetic integral control of stochastic reaction networks

The antithetic integral controller has been introduced in [15] with the aim of developing an integral control theory for stochastic biochemical networks. Integral control is a cornerstone of control theory and control 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 represented as a set of chemical reactions in order to be implemented in-vivo, such as in bacteria [4]. This controller takes the form of a (stochastic) reaction network $(\boldsymbol{X} \cup \boldsymbol{Z}, \mathcal{R}_{\text{AIC}})$ with species $\boldsymbol{Z} := \{\boldsymbol{Z_1}, \boldsymbol{Z_2}\}$ and reactions

$$\boldsymbol{Z_1} \xrightarrow{k} \boldsymbol{Z_1} + \boldsymbol{X_1}, \boldsymbol{X_\ell} \xrightarrow{\theta} \boldsymbol{X_\ell} + \boldsymbol{Z_2}, \emptyset \xrightarrow{\mu} \boldsymbol{Z_1}, \boldsymbol{Z_2} + \boldsymbol{Z_1} \xrightarrow{\eta} \emptyset \tag{11}$$

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

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

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

We have the following result in the case of unimolecular networks with mass-action kinetics which is a slight extension of the one in [15]:

Proposition 8 *Let us consider the stochastic reaction network $(\boldsymbol{X}, \mathcal{R})$ with first-order mass-action kinetics satisfying Assumption 3. Then, the following statements are equivalent:*

(a) there exist vectors $v \in \mathbb{R}^d_{>0}$ and $w \in \mathbb{R}^d_{\geq 0}$, $w_1, w_\ell > 0$,

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

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

(b) 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[1]*.*

(c) The closed-loop network $(\boldsymbol{X} \cup \boldsymbol{Z}, \mathcal{R}_{\text{AIC}})$ is ergodic and we have that

$$\mathbb{E}[X_\ell(t)] \to \mu/\theta \text{ as } t \to \infty \tag{13}$$

Proof : The proof of the equivalence between the statement (a) and (b) has been obtained in [15] as well as the proof of the implication (a) $\Rightarrow$ (c). To prove the implication (c) $\Rightarrow$ (b) it is enough to remark that the stability and the output controllability of the first -order moment equation are necessary conditions allowing for the integral control of a system. Indeed, if the moment equation is unstable, then so will be the first-order moments of the closed-loop network and if the system is not output controllable, then one cannot change the value of output by suitably acting on the input. This completes the proof. $\diamondsuit$

As in Theorem 4, the above result connects algebraic conditions to the stability and the output-controllability of a linear time-invariant system, which are standard concepts of systems and control theory. The algebraic conditions can be solved using standard linear programming techniques. However, 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 in the deterministic setting often results into the destabilization of the closed-loop dynamics (unless the system to be controlled satisfies certain strong properties).

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

[1] i.e. rank $\begin{bmatrix} e_\ell^T e_1 & e_\ell^T S_u W_u e_1 & \dots & e_\ell^T (S_u W_u)^{d-1} e_1 \end{bmatrix} = 1$

Michaelis-Menten kinetics [15]. However, the existing theoretical results do not yet cover those important classes of reaction networks. The difficulty arises from the potential loss of the (global) output-controllability condition. Simulation results tend to suggest that this controller should work properly even when the output-controllability property is not met globally, as is often the case for 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 admit a natural reaction network representation and are relevant to consider. We first recall some theoretical basics on phase-type distributions and provide some examples in order to illustrate their richness in terms of behavior. In fact, those distributions are dense in the set of all probability distributions, which means that we can approximate arbitrarily closely any distribution, including heavy-tailed ones. We then provide a complete characterization of phase-type distributions in terms of simple algebraic conditions. Finally, under the existence assumption of a Markov process describing the considered phase-type distribution, we propose a simple procedure to construct the unique minimal unimolecular reaction network encoding this distribution. By unique and minimal, it is meant here that it is both the only network that exactly represents the considered distribution and the one that has the smallest number of reactions and molecular species.

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 (also known as phases) placed in series and parallel. It is represented by a random variable describing the time until absorption of a Markov process with one absorbing state starting from the 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

$$f(\tau) = \alpha e^{H\tau} H_0, \ \tau \geq 0 \tag{14}$$

where $\alpha \in \mathbb{R}^{1\times m}_{\geq 0}$, $||\alpha||_1 = 1$, $H \in \mathbb{R}^{m\times m}$ is a Hurwitz stable Metzler matrix and $H_0 := -H\mathbb{1}_n$. The parameter H is referred here to as the *subgenerator matrix* and α as the *input probability row vector*. Let $X \sim \mathrm{PH}(\alpha, H)$, then all the moments of this random variable are given by

$$\mathbb{E}[X^\ell] = (-1)^\ell \ell! \alpha H^{-\ell} \mathbb{1}_m. \tag{15}$$

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

$$\dot{p}(t) = p(t) \begin{bmatrix} 0 & 0_{1\times m} \\ H_0 & H \end{bmatrix} \text{ with } p(0) = \begin{bmatrix} 0 & \alpha \end{bmatrix}. \tag{16}$$

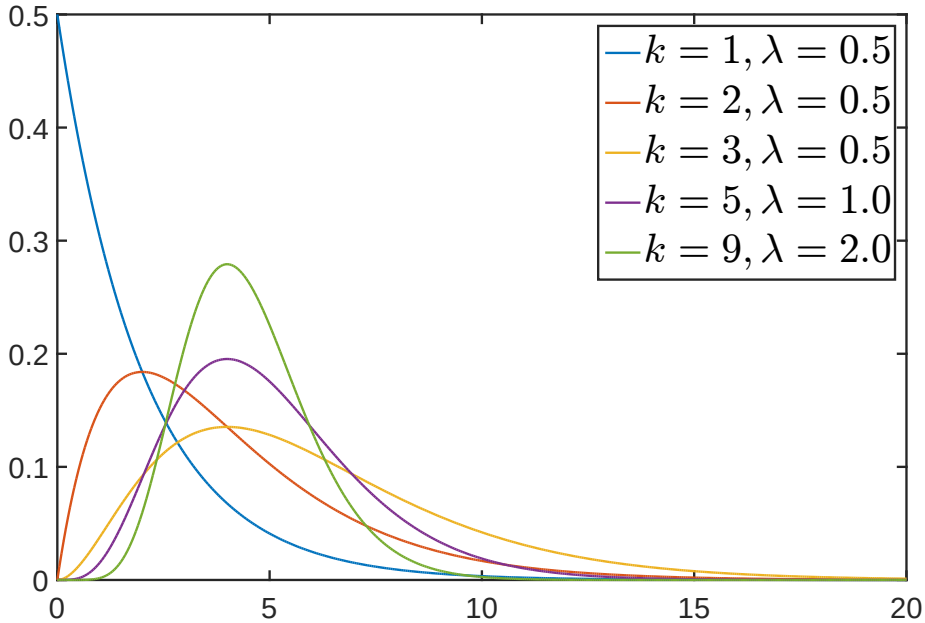

Figure 1: Examples of Erlang distributions

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 quite rich as it can take various forms. In the case of a hypoexponential distribution with four phases and four parameters $\lambda_1, \ldots, \lambda_4 > 0$, the matrix H is given by

$$H = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 \\ 0 & 0 & 0 & -\lambda_4 \end{bmatrix}. \tag{17}$$

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

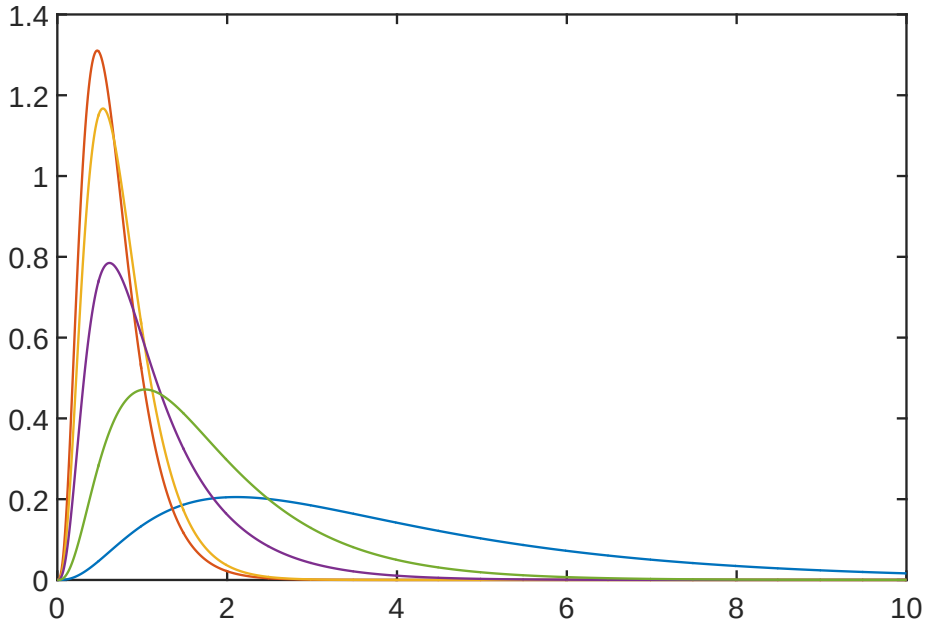

Figure 2: Examples of hypoexponential distributions

Figure 3 consists of the average of the density functions associated with the aforementioned Erlang distributions.

3.2 Approximation of a constant delay

In the deterministic setting, constant delays can be approximated arbitrarily closely by sequence of filters such as low-pass filters [26,47] or all-pass filters such as lattice networks or Padé approximants [44,47,60]. It is 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 series of N low-pass filters with overall transfer function given by

$$H_N(s) := \left(1 + \frac{\bar{\tau}s}{N}\right)^{-N}, N \in \mathbb{Z}_{>0}. \tag{18}$$

Note, moreover, that $e^{-\bar{\tau}s}$ is the Laplace transform of the Dirac distribution $\delta(t-\bar{\tau})$ centered around $\bar{\tau}$. We have the following standard result which is recalled for completeness:

Proposition 9 ([26, 47]) *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,

$$\begin{aligned} H_N(s) &= \exp(\log(H_N(s))) \\ &= \exp(-N\log(1+\bar{\tau}s/N)) \\ &\simeq \exp(-\bar{\tau}s) \end{aligned} \tag{19}$$

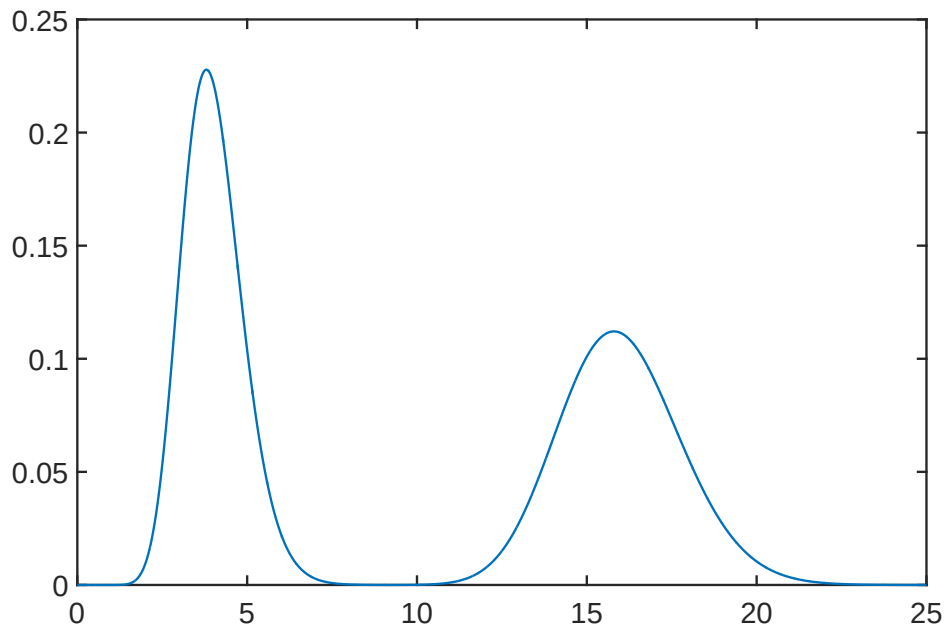

Figure 3: Example of a bimodal Hyper-Erlang distribution.

where we have used the principal branch of the logarithm and the fact that $\log(1+\bar{\tau}s/N) \approx \bar{\tau}s/N$ for a sufficient large N. As a result, 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 (in distribution) of an Erlang random variable. This is stated in the following result:

Proposition 10 *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*

$$f_N(\xi) = \frac{N^N \xi^{N-1} e^{-N\xi/\bar{\tau}}}{\bar{\tau}^N (N-1)!}. \tag{20}$$

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

Proof : Instead of proving that the cumulative distribution $F_N(\xi)$ of τ_N converges to the shifted Heaviside function $H(\xi-\bar{\tau})$ at all continuity points of $H(\xi-\bar{\tau})$ 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 (15), it is immediate to see that the mean of τ_N is given by

$$\mathbb{E}[\tau_N] = \bar{\tau} \tag{21}$$

and its variance by

$$V(\tau_N) = \bar{\tau}^2/N. \tag{22}$$

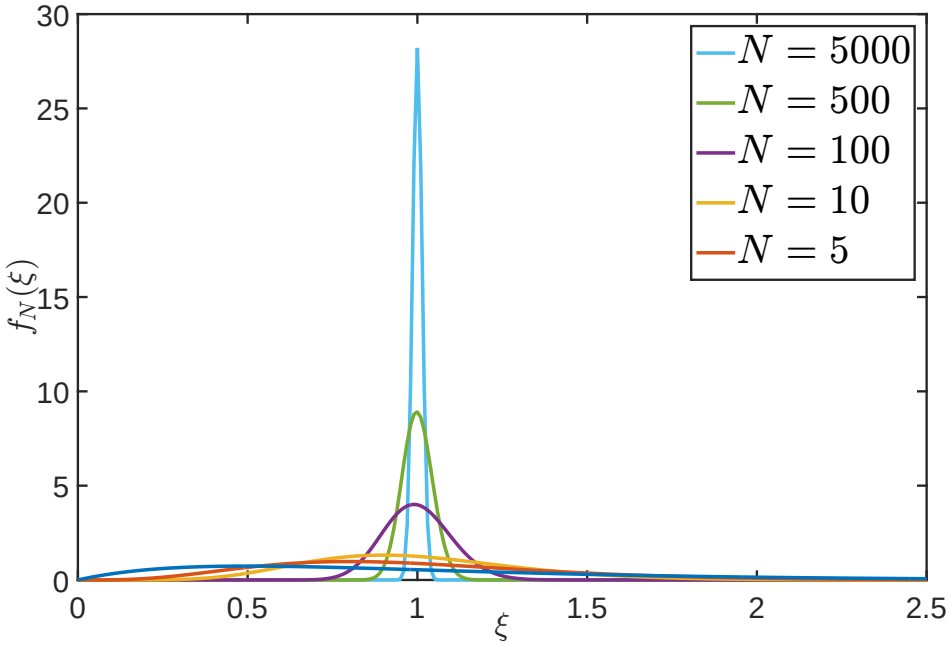

Figure 4: Erlang distributions with rate $N/\bar{\tau}$ and shape N approximating the Dirac distribution.

Hence, we have that $\mathbb{E}[\tau_N] \to \bar{\tau}$ and $V(\tau_N) \to 0$ as $N \to \infty$. Using the fact that the Erlang distributions are unimodal, then this implies that the density function $f_N(\xi)$ converges to the shifted Dirac delta $\delta(\xi - \bar{\tau})$, and hence that $F_N(\xi)$ of τ_N converge to the shifted Heaviside function $H(\xi - \bar{\tau})$ as $N \to \infty$. 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.

3.3 Reaction network implementation of phase-type delays

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

3.3.1 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 11 *Let us consider a matrix $H \in \mathbb{R}^{m\times m}$.*

(a) The matrix $H \in \mathbb{R}^{m\times m}$ is a subgenerator matrix.

(b) The following conditions hold:

(i) it is Metzler and Hurwitz stable,

(ii) $H\mathbb{1}_m \leq 0$, and

(iii) at least one entry of $H\mathbb{1}_n$ is nonzero; i.e. $H\mathbb{1}_m \neq 0$.

Proof : The matrix H is a subgenerator matrix if and only if the matrix

$$\bar{H} := \begin{bmatrix} 0 & 0 \\ -H\mathbb{1} & H \end{bmatrix} \tag{23}$$

is a Q-matrix (i.e. the matrix $\bar{H}$ is Metzler and $\bar{H}\mathbb{1}_m = 0$) and that the first state is absorbing. Clearly, a necessary and sufficient condition for $\bar{H}$ to be Metzler is that H be Metzler such that $H\mathbb{1} \leq 0$. The last condition is obtained because the first state needs to be absorbing. This is equivalent to saying that $H\mathbb{1}_n \neq 0$. This proves the result. $\diamondsuit$

3.3.2 Construction of the reaction network

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

- 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 needs to be irreducible.

The first constraint is easily met by exclusively considering 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-absorption 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. In this regard, adding delay lines should not cause of a loss of irreducibility.

The following result describes the construction of the unique minimal stochastic reaction network implementing a given phase-type distributed delay:

Proposition 12 *Let the matrix $H = [h_{ij}]_{i,j=1,\dots,m}$ be a subgenerator matrix and $\alpha \in \mathbb{R}^{1\times n}$ be an input probability row vector. Then, the delay $\tau \sim \mathrm{PH}(\alpha, H)$ is exactly represented by the stochastic reaction network*

$$\emptyset \xrightarrow{\alpha e_i} \boldsymbol{D_i}, \quad \boldsymbol{D_i} \xrightarrow{h_{ij}} \boldsymbol{D_j}, \quad \boldsymbol{D_i} \xrightarrow{-e_i^T H \mathbb{1}} \emptyset, \tag{24}$$

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

Proof : The queue (consisting of the reactions involving delay species as both reactants and products) is conservative since it only consists of conversion reactions. From the structure of the matrix H, 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 needed and one cannot choose a smaller number. This proves the minimality of the number of reactions and, as a consequence, this network is unique. ♢

The above result shows that any phase-type distributed delay can be represented as a simple unimolecular chemical reaction network. Each produced molecule by the birth reactions gets destroyed by the death reactions after a delay $\tau \sim \mathrm{PH}(H, \alpha)$. This is very convenient since it is known from [17, 18, 32] that unimolecular stochastic reaction networks are usually well-behaved and easy to analyze using standard linear algebra tools.

Example 13 *Let us consider a delay that follows the Erlang distribution with rate $\lambda > 0$ and shape m. In this case, we have that*

$$H = \begin{bmatrix} -\lambda & \lambda & 0 & \dots & 0 \\ 0 & -\lambda & \lambda & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & -\lambda \end{bmatrix}, H_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ \lambda \end{bmatrix}, \alpha = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}^T. \tag{25}$$

Elementary matrix calculations show that the probability density function of the time to absorption $\alpha e^{H\xi} H_0$ is equal to

$$\frac{\lambda^N \xi^{N-1} e^{-\lambda\xi}}{(N-1)!}, \tag{26}$$

which confirms that $\mathrm{PH}(H, \alpha)$ generates an Erlang distribution with shape N and rate λ. The corresponding reaction network is given by

$$\emptyset \xrightarrow{1} \boldsymbol{D_1}, \quad \boldsymbol{D_i} \xrightarrow{\lambda} \boldsymbol{D_{i+1}}, \quad \boldsymbol{D_m} \xrightarrow{\lambda} \emptyset, i = 1, \dots, m-1. \tag{27}$$

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

$$f_1(\boldsymbol{X}) + f_2(\boldsymbol{X}) \xrightarrow{k,\tau(H,\alpha)} f_1(\boldsymbol{X}) + g(\boldsymbol{X}) \tag{28}$$

where $f_1(\boldsymbol{X})$ and $f_2(\boldsymbol{X})$ denote the preserved part of the reactants and the consumed part of the reactants, respectively. The term $g(\boldsymbol{X})$ represents the produced species/reactants. 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 $\tau \sim \mathrm{PH}(H,\alpha)$. We can then apply the construction procedure of Section 3.3 to get the delay line. However, we now interconnect the substitute the reactants of the birth reactions in the delay line by the reactant of the delayed reaction (28) and the products of the death reactions of the delay line by the products of the reaction (28). However, attention should be paid as the reaction (28) can be interpreted in two different ways depending on its meaning. This is formalized below:

Proposition 14 *The delayed reaction* (28) *with the m-phase delay* $\tau \sim \mathrm{PH}(H,\alpha)$ *admits one of the following realizations:*

(a) Non-absorbing realization:

$$\begin{aligned} f_1(\boldsymbol{X}) + f_2(\boldsymbol{X}) &\xrightarrow{k\alpha e_i} f_1(\boldsymbol{X}) + \boldsymbol{D_i}, \\ \boldsymbol{D_i} &\xrightarrow{h_{ij}} \boldsymbol{D_j}, \\ \boldsymbol{D_i} &\xrightarrow{-e_i^T H \mathbb{1}_m} g(\boldsymbol{X}) \end{aligned} \tag{29}$$

for all $i = 1, \ldots, m$, $i \neq j$.

(b) Absorbing realization:

$$\begin{aligned} f_1(\boldsymbol{X}) + f_2(\boldsymbol{X}) &\xrightarrow{k\alpha e_i} \boldsymbol{D_i}, \\ \boldsymbol{D_i} &\xrightarrow{h_{ij}} \boldsymbol{D_j}, \\ \boldsymbol{D_i} &\xrightarrow{-e_i^T H \mathbb{1}_m} f_1(\boldsymbol{X}) + g(\boldsymbol{X}) \end{aligned} \tag{30}$$

for all $i, j = 1, \ldots, m$, $i \neq j$.

Proof : It is immediate to see that we have two possible implementations for this delayed reaction. Either the preserved part of the reaction is non-absorbed in the queue or it is absorbed and restituted at the end of the queue. ♢

We justify the above result with the following examples:

Example 15 *Let us consider for instance the catalytic reaction*

$$\boldsymbol{X_1} + \boldsymbol{X_2} \xrightarrow{k,\tau(H,\alpha)} \boldsymbol{X_1} + \boldsymbol{X_2} + \boldsymbol{X_3} \tag{31}$$

where $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$ represent mRNA, ribosomes and protein molecules, respectively. Let us assume that the delay represents the protein maturation after translation, the letter being assumed here to happen instantaneously. In this regard, the mRNA and ribosome molecules should not be absorbed in the queue and we have

$$\begin{aligned} \boldsymbol{X_1} + \boldsymbol{X_2} &\xrightarrow{k\alpha e_i} \boldsymbol{X_1} + \boldsymbol{X_2} + \boldsymbol{D_i}, \\ \boldsymbol{D_i} &\xrightarrow{h_{ij}} \boldsymbol{D_j}, \\ \boldsymbol{D_i} &\xrightarrow{-e_i^T H \mathbb{1}_m} \boldsymbol{X_3} \end{aligned} \tag{32}$$

where $, i, j = 1, \ldots, m, i \neq j$.

Example 16 *Let us consider the same reaction as in the previous example but assume now that the delay represents the translation delay. In this regard, the ribosome and mRNA molecules should be absorbed in the queues since these molecules are made unavailable when the translation reaction takes place. Therefore, the queue should be made absorbing as*

$$\begin{aligned} \boldsymbol{X_1} + \boldsymbol{X_2} &\xrightarrow{k\alpha e_i} \boldsymbol{D_i}, \\ \boldsymbol{D_i} &\xrightarrow{h_{ij}} \boldsymbol{D_j}, \\ \boldsymbol{D_i} &\xrightarrow{-e_i^T H \mathbb{1}_m} \boldsymbol{X_1} + \boldsymbol{X_2} + \boldsymbol{X_3} \end{aligned} \tag{33}$$

where $, i, j = 1, \ldots, m, i \neq j$.

The above implementations may not be completely equivalent from the analysis viewpoint. Indeed, the first one may be more difficult to consider because of the catalytic nature of the bimolecular reaction which only produces mass. The second implementation may be simpler to consider as it involves a conversion reaction that both destroys and produces mass.

We are now in position to define the networks of interest:

Definition 17 (Delayed reaction network) *A delayed reaction network is defined as the triplet $(\boldsymbol{X}, \mathcal{R}, \tau)$ where $\boldsymbol{X}$ is the set of molecular species, $\mathcal{R}$ is the set of reactions and τ is the vector of delays.*

Definition 18 (Delay-free reaction network) *Let us consider a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$. Its associated delay-free network is defined as $(\boldsymbol{X}, \mathcal{R}, 0) = (\boldsymbol{X}, \mathcal{R})$ and consists of setting all the delays to zero.*

Definition 19 (Augmented reaction network) *Let us consider a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ and assume that the delays are phase-type distributed. Let us further decompose $\mathcal{R}$ as two disjoint sets $\mathcal{R}^0$ and $\mathcal{R}^\tau$ containing the non-delayed and the delayed reactions, respectively. Then, its associated augmented network is defined as $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup, \mathcal{R}_d)$ where $\boldsymbol{D}$ and $\mathcal{R}_d$ are the set of delay species and delay reactions, respectively.*

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 20 *Let us consider a delayed reaction network* $(\boldsymbol{X}, \mathcal{R}, \tau)$ *with phase-type distributed delays. Then, the following statements are equivalent:*

(a) The projected state-space onto $\mathbb{Z}^d_{\geq 0}$ *of the Markov process associated with the network* $(\boldsymbol{X}, \mathcal{R}, \tau)$ *is irreducible.*

(b) The state-space of the Markov process associated with the augmented network $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup, \mathcal{R}_d)$ *is irreducible.*

(c) The state-space of the Markov process associated with the delay-free network $(\boldsymbol{X}, \mathcal{R})$ *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-space by construction, hence the irreducibility only depends on the reaction network with the delays set to 0. This proves the equivalence between the two last statements. To prove the implication (c) $\Rightarrow$ (a), just note that delays are only temporal features and cannot change the irreducibility property of the state-space projected onto $\mathbb{Z}^d_{\geq 0}$ provided that the delays are bounded with probability one, which is the case for phase-type distributed delays. The reverse implication is immediate here. $\diamondsuit$

5.2 Ergodicity of unimolecular networks with delays

Let us consider here the delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with d species, K reactions and n delays. Each delay τ is phase-type distributed, i.e. $\tau_k \sim \mathrm{PH}(H_i, \alpha_i)$, $i = 1, \ldots, n$, for some subgenerator matrix $H_i \in \mathbb{R}^{d_i \times d_i}$ and some input probability row vector $\alpha_i \in \mathbb{R}^{d_i}_{\geq 0}$. We can then define the augmented network $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup, \mathcal{R}_d)$. The stoichiometric matrix S associated with this augmented reaction network is given by

$$H = \left[\begin{array}{c|c|c|c} S_x & S_{\mathrm{in},x} & S_{\mathrm{out},x} & 0 \\ \hline 0 & S_{\mathrm{in},d} & S_{\mathrm{out},d} & S_d \end{array}\right] \tag{34}$$

where S_x is the stoichiometric matrix associated with the reactions that do not involve any delay species and, conversely, S_d is the stoichiometric matrix associated with reactions that only involve delay species. The stoichiometric matrices

$$S_{\mathrm{in}} := \begin{bmatrix} S_{\mathrm{in},x} \\ S_{\mathrm{in},d} \end{bmatrix} \text{ and } S_{\mathrm{out}} := \begin{bmatrix} S_{\mathrm{out},x} \\ S_{\mathrm{out},d} \end{bmatrix} \tag{35}$$

corresponds to conversion reactions entering the delay lines and leaving the delay lines, respectively. The associated propensity functions are given by

$$\lambda_x(x) := \begin{bmatrix} W_x x \\ b_x \end{bmatrix}, \lambda_{\text{in}}(x) := \begin{bmatrix} W_{\text{in}} x \\ b_{\text{in}} \end{bmatrix}, \lambda_{\text{out}}(\delta) := W_{\text{out}}\delta \text{ and } \lambda_d(\delta) := W_d\delta \tag{36}$$

where $(x, \delta) \in \mathbb{Z}^d_{\geq 0} \times \mathbb{Z}^{d_1+\ldots+d_n}_{\geq 0}$ is the state of the augmented network. The terms $W_x x$, b_x, $W_{\text{in}} x$, b_{in}, $W_{\text{out}}\delta$ and $W_d\delta$ represent the propensity functions of the non-delayed first-order reactions, the non-delayed zeroth-order reactions, the first-order reactions entering the delay lines, the zeroth-order reactions entering the delay lines, the propensity of the reactions leaving the delay lines and the propensity of the reactions inside the delay lines, respectively.

Let us also define the following matrices

$$\begin{aligned} A &:= S_x \begin{bmatrix} W_x \\ 0 \end{bmatrix} + S_{\text{in},x} \begin{bmatrix} W_{\text{in}} \\ 0 \end{bmatrix} \\ B &:= S_{\text{out},x} W_{\text{out}} \\ C &:= S_{\text{in},d} \begin{bmatrix} W_{\text{in}} \\ 0 \end{bmatrix} \\ D &:= S_{\text{out},d} W_{\text{out}} + S_d W_d \end{aligned} \tag{37}$$

Calculations show that

$$\begin{aligned} W_{\text{out}} &:= -\operatorname*{diag}_{i=1}^{n}(\mathbb{1}^T H_i^T) \\ S_{\text{in},d} &= \operatorname*{diag}_{i=1}^{n}(\alpha_i^T) \\ W_{\text{in}} &= \begin{bmatrix} q_1^T \\ \vdots \\ q_m^T \end{bmatrix} \\ C &= \left[\begin{array}{c} \alpha_1^T q_1^T \\ \vdots \\ \alpha_m q_n^T \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \\ D &= H^T = \operatorname*{diag}_i(H_i)^T \end{aligned} \tag{38}$$

where $q_j = r_j e_{\sigma(j)}$ and 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 $\boldsymbol{X_i}$ acts on the production of some delay species in the delay line j.

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

$$S_{\text{df}} := \left[S_x \;\middle|\; S_{\text{in},x} + S_{\text{out},x} \right] \tag{39}$$

with the propensity function

$$\lambda_{\mathrm{df}}(x) := \left[\begin{array}{c} W_x x \\ b_x \\ \hline W_{\mathrm{in}} x \\ b_{\mathrm{in}} \end{array}\right] \tag{40}$$

which leads to the characteristic matrix

$$A_{\mathrm{df}} = S_x \begin{bmatrix} W_x \\ 0 \end{bmatrix} + (S_{\mathrm{in},x} + S_{\mathrm{out},x}) \begin{bmatrix} W_{\mathrm{in}} \\ 0 \end{bmatrix}. \tag{41}$$

The following proposition will be instrumental in proving the main result of the section:

Proposition 21 *We have $A_{\mathrm{df}} = A - BH^{-T}C$.*

Proof : Comparing the expressions for A_{df} and A, we simply need to prove that

$$-W_{\mathrm{out}} H^{-T} C = \begin{bmatrix} W_{\mathrm{in}} \\ 0 \end{bmatrix}. \tag{42}$$

We know that $W_{\mathrm{in}} = \underset{i=1}{\overset{n}{\mathrm{col}}}(q_i^T)$. Now, we have

$$\begin{aligned} -W_{\mathrm{out}} H^{-T} C &= \underset{i=1}{\overset{n}{\mathrm{diag}}}(\mathbb{1}^T H_i^T) H^{-T} \left[\begin{array}{c} \underset{i=1}{\overset{n}{\mathrm{col}}}(\alpha_i^T q_i^T) \\ \hline 0 \end{array}\right] = \underset{i=1}{\overset{n}{\mathrm{diag}}}(\mathbb{1}^T) \left[\begin{array}{c} \underset{i=1}{\overset{n}{\mathrm{col}}}(\alpha_i^T q_i^T) \\ \hline 0 \end{array}\right] \\ &= \left[\begin{array}{c} \underset{i=1}{\overset{n}{\mathrm{col}}}(q_i^T) \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} W_{\mathrm{in}} \\ \hline 0 \end{array}\right]. \end{aligned} \tag{43}$$

This completes the proof. ♢

We can now state our main result on the ergodicity of delayed unimolecular reaction networks:

Theorem 22 *Let us consider a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with zeroth- and firth-order mass-action kinetics and phase-type distributed delays for which the delay-free counterpart $(\boldsymbol{X}, \mathcal{R})$ satisfies Assumption 3. Then, the following statements are equivalent:*

(a) The delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ is exponentially ergodic.

(b) The augmented reaction network $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup, \mathcal{R}_d)$ is exponentially ergodic.

(c) The matrix

$$\mathcal{A} := \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right]$$

is Hurwitz stable.

(d) The matrix A_{df} is Hurwitz stable.

(e) The delay-free network $(\boldsymbol{X}, \mathcal{R})$ is exponentially ergodic.

Proof : By assumption, the state-space of the Markov process associated with the delay-free reaction network is irreducible, Equivalently, the state-spaces of the Markov processes associated with the augmented and the delayed reaction networks are also irreducible. Since, the delay-free as no closed components. This is equivalent to saying that the augmented and the delayed reaction networks do not have closed components as well.

The statements (a) and (b) are equivalent due to the equivalence of representations.

To prove the equivalence between (b) and (c), let $\tilde{\mathbb{A}}$ be the generator of Markov process representing the augmented reaction network and define the linear form $V(x,\delta) := v_x^T x + v_\delta^T \delta$ defined for $v_x \in \mathbb{R}^d_{>0}$ and $v_\delta \in \mathbb{R}^{d_1+\ldots+d_n}_{>0}$. Then, the Markov process representing the augmented network 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 $(x,\delta) \in \mathbb{Z}^d_{\geq 0} \times \mathbb{Z}^{d_1+\ldots+d_n}_{\geq 0}$. This can be reformulated as

$$\tilde{\mathbb{A}}\left(v_x^T x + v_\delta^T \delta\right) = \begin{bmatrix} v_x \\ v_\delta \end{bmatrix}^T \left(\left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] \begin{bmatrix} x \\ \delta \end{bmatrix} + \left[\begin{array}{c} S_x b_x \\ \hline S_{\mathrm{in},d} b_{\mathrm{in}} \end{array}\right] \right). \tag{44}$$

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 $(x,\delta) \in \mathbb{Z}^d_{\geq 0} \times \mathbb{Z}^{d_1+\ldots+d_n}_{\geq 0}$ if and only if $\mathcal{A}$ is Hurwitz stable. This proves the equivalence between the statements (b) and (c).

To prove the equivalence between the statements (c), (d) and (e), we first need the following result:

Lemma 23 *Let us consider a Metzler matrix $M = [M_{ij}]_{i,j=1}^2$ where both M_{11} and M_{22} are square. Then, the matrix M is Hurwitz stable if and only if the matrices $M_{11} - M_{12} M_{22}^{-1} M_{21}$ and M_{22} are Hurwitz stable.*

Since H is Hurwitz stable by construction, we can apply the above result with $M_{11} = A$, $M_{12} = B$, $M_{21} = C$ and $M_{22} = H^T$. As a result we get that $\mathcal{A}$ is Hurwitz stable if and only if $A - BH^{-T}C$ is also Hurwitz stable. From Proposition 21, this is equivalent to say that A_{df} is Hurwitz stable and that the delay-free network is exponentially ergodic. This proves the result. $\diamondsuit$

The above result is interesting for multiple reasons. The first one is that the shape of the distribution or, equivalently, its complexity does not have any impact of the ergodicity of the network as long as it is unimolecular. A conclusion is that one can consider *almost constant deterministic delays* as one can choose an arbitrarily large, yet finite, shape value in the Erlang distribution. Yet, one cannot conclude on the ergodicity of the network in the case of a constant deterministic delay.

5.3 Ergodicity of bimolecular networks with delays

We address here the case of bimolecular networks with delays and we will distinguish two cases. The first one is when only unimolecular networks are delayed whereas the second one is when the network is allowed to have delayed bimolecular reactions.

5.3.1 No bimolecular reactions enters the queue

We consider here a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with zeroth-, first- and second-order mass-action kinetics and phase-type distributed delays. We can define S_b to be the stoichiometric matrix of the augmented network associated with the bimolecular reactions as

$$S_b = \left[\begin{array}{c} S_{b,x} \\ \hline 0 \end{array}\right]. \tag{45}$$

The rest of the matrices are the same as for unimolecular networks; see Section 5.2. We then have the following result:

Theorem 24 *Let us consider a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with zeroth-, first- and second-order mass-action kinetics and phase-type distributed delays for which the delay-free counterpart $(\boldsymbol{X}, \mathcal{R})$ satisfies Assumption 3. Then, the following statements are equivalent:*

(a) There exist vectors $v_x \in \mathbb{R}^d_{>0}$ and $v_\delta \in \mathbb{R}^{d_1+\ldots+d_n}_{>0}$ verifying $\left[v_x^T \;\middle|\; v_\delta^T\right] S_b = 0$ for which the function $V(x,\delta) = v_x^T x + v_\delta^T \delta$ is a Foster-Lyapunov function for the Markov process.

(b) There exist vectors $v_x \in \mathbb{R}^d_{>0}$ and $v_\delta \in \mathbb{R}^{d_1+\ldots+d_n}_{>0}$ such that the conditions

$$\left[\begin{array}{c} v_x \\ \hline v_\delta \end{array}\right]^T \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] < 0 \text{ and } \left[v^T \;\middle|\; v_\delta^T\right] S_b = 0$$

hold.

(c) There exist a vector $v \in \mathbb{R}^d_{>0}$ satisfying $v^T S_{b,x} = 0$ and $v^T A_{\mathrm{df}} < 0$.

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 : The equivalence from the two first statements comes from a direct application of Theorem 2. The equivalence with the last one follows from the same arguments as for Theorem 22. ♢

5.3.2 At least one bimolecular reactions enters the queue

We consider here a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with zeroth-, first- and second-order mass-action kinetics and phase-type distributed delays. As the case of delayed unimolecular reactions can be brought back to non-delayed unimolecular reactions, we assume here that the delays only act on bimolecular reactions.

The stoichiometric matrix of the delayed-network associated with the non-delayed bimolecular reactions is denoted by S_b whereas

$$S_b^d = \begin{bmatrix} \zeta_{r,1}^d - \zeta_{\ell,1}^d & \dots & \zeta_{r,K_d}^d - \zeta_{\ell,K_d}^d \end{bmatrix} \tag{46}$$

is the stoichiometric matrix associated with the delayed bimolecular reactions where K_d is the number of delayed biomolecular reactions. After augmenting the network to incorporate the delay species and reactions, the latter becomes

$$S_b^i = \begin{bmatrix} -\zeta_{\ell,1}^d \mathbb{1}_{\varsigma_1}^T & \dots & -\zeta_{\ell,K}^d \mathbb{1}_{\varsigma_K}^T \\ J_i^1 & & \\ & \ddots & \\ & & J_i^K \end{bmatrix} \tag{47}$$

where J_i^k is a matrix with entries equal to zero or one, $k = 1, \dots, K_d$, such that on each column there is one and only one entry equal to one. We also define the matrix

$$B_b = \begin{bmatrix} J_1^o & \dots & J_K^o \end{bmatrix} \tag{48}$$

where $J_k^o = \zeta_{r,k}^d \mathbb{1}^T \Lambda_k^o$ and Λ_k^o is a diagonal matrix with nonnegative entries representing the rate parameters of the reactions leaving the queue.

We then have the following result:

Theorem 25 *Let us consider a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with zeroth-, first- and second-order mass-action kinetics and phase-type distributed delays. for which the delay-free counterpart $(\boldsymbol{X}, \mathcal{R})$ satisfies Assumption 3. Then, the following statements are equivalent:*

(a) There exist $v_x \in \mathbb{R}_{>0}^d$ and $v_\delta \in \mathbb{R}^{d_1+\dots+d_n}$, verifying $v_x^T S_b = 0$ and $\begin{bmatrix} v_x^T & v_\delta^T \end{bmatrix} S_b^i = 0$ such that the function $V(x, \delta) = v_x^T x + v_\delta^T \delta$ is a Foster-Lyapunov function for the associated Markov process.

(b) There exist $v_x \in \mathbb{R}_{>0}^d$ and $v_\delta \in \mathbb{R}^{d_1+\dots+d_n}$, such that the conditions

$$v^T S_b = 0, \quad \begin{bmatrix} v_x^T & \mid & v_\delta^T \end{bmatrix} S_b^i = 0 \tag{49}$$

and

$$\begin{bmatrix} v \\ \hline v_\delta \end{bmatrix} \left[\begin{array}{c|c} A & B_b \\ \hline 0 & H^T \end{array}\right] < 0 \tag{50}$$

where $H = \mathrm{diag}_{i=1}^{K_d}(H_i)$ and A is the same as for unimolecular networks.

(c) There exists a $v \in \mathbb{R}^d_{>0}$ such that

(i) $v^T A < 0$,

(ii) $v^T S_b = 0$, *and*

(iii) $v^T S_b^d < 0$.

Moreover, when one of the above statements hold, then the Markov process describing the augmented reaction network $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup \mathcal{R}_d)$ *is exponentially ergodic and its stationary distribution is light-tailed.*

Proof : Similar calculations as in the previous proofs show that the conditions of the first statement can be equivalently reformulated as the conditions in statement (b). This proves the equivalence between the statements (a) and (b). The conditions of statement (b) can be reformulated as

$$\begin{array}{ccc} v_x^T S_b & = & 0 \\ -v_x^T \zeta_{\ell,1}^d \mathbb{1}^T + v_{\delta,1}^T J_i^1 & = & 0 \\ \vdots & \vdots & \vdots \\ -v_x^T \zeta_{\ell,K_d}^d \mathbb{1}^T + v_{\delta,K_d}^T J_i^{K_d} & = & 0 \\ v_x^T A & < & 0 \\ v_x^T J_1^0 + v_{\delta,1}^T H_1^T & < & 0 \\ \vdots & \vdots & \vdots \\ v_x^T J_{K_d}^0 + v_{\delta,K_d}^T H_{K_d}^T & < & 0, \end{array} \tag{51}$$

where $v_\delta =: (v_{\delta,1}, \ldots, v_{\delta,K_d}) \in \mathbb{R}^{d_1}_{>0} \times \mathbb{R}^{d_{K_d}}_{>0}$. We can put aside the conditions $v_x^T A < 0$ and $v_x^T S_b = 0$ (two first conditions in the statement (c)) and consider each of the $v_{\delta,k}$'s separately as the corresponding conditions are uncoupled. The idea is to turn the system of inequality and equality conditions into a system of equality conditions. To this aim, let us define a positive vector q_k of appropriate dimensions and consider the equality $v_x^T J_k^0 + v_{\delta,k}^T H_k^T = -q_k^T$. Since the matrix H_k is Metzler and Hurwitz stable by construction, then we have that $H_k^{-1} \leq 0$. Hence, we have that $v_x^T J_k^0 H_k^{-T} + v_{\delta,k}^T = -q_k^T H_k^{-T}$ or, equivalently, $v_{\delta,k}^T = -v_x^T J_k^0 H_k^{-T} - q_k^T H_k^{-T}$. Combining this with the other equality yields

$$\left[v_x^T J_k^0 H_k^{-T} + q_k^T H_k^{-T}\right] J_i^k + v_x^T \zeta_{\ell,k}^d \mathbb{1}_{\varsigma_k} = 0 \tag{52}$$

We note that

$$\begin{array}{rcl} J_k^0 H_k^{-T} J_i^k & = & \zeta_{\ell,k}^d \mathbb{1}^T \Lambda_k^o H_k^{-T} J_i^k \\ & = & -v_x^T \zeta_{r,k}^d \mathbb{1}_{\varsigma_k}^T \end{array} \tag{53}$$

where we have used the fact that $\mathbb{1}^T \Lambda_k^o H_k^{-T} J_i^k$ is the negative of the static-gain of the delay-line, which is equal to a vector of ones due to the property of conservativity of delays. This leads to the following equality

$$v_x^T (\zeta_{\ell,k}^d - \zeta_{r,k}^d) \mathbb{1}_{\varsigma_k} = -q_k^T H_k^{-T} J_i^k, \tag{54}$$

for $k = 1, \ldots, K_d$. Since $q_k > 0$, then this means that a necessary condition for the conditions in (51) is that $v_x^T(\zeta_{r,k} - \zeta_{\ell,k}) < 0$. We know that $H_k^{-T} J_i^k$ is full column rank since H_k is invertible and J_i^k is full-column rank. In this regard for any given v_x, we can find a q_k such that (54) holds. This proves the equivalence with the last statement. $\diamondsuit$

Example 26 *Let us consider the epidemiological network*

$$\begin{aligned}
\emptyset &\xrightarrow{k_1} \boldsymbol{X_1} \xrightarrow{\gamma_1} \emptyset, \\
\emptyset &\xrightarrow{k_2} \boldsymbol{X_2} \xrightarrow{\gamma_2} \emptyset, \\
\emptyset &\xrightarrow{k_3} \boldsymbol{X_3} \xrightarrow{\gamma_3} \emptyset, \\
\boldsymbol{X_2} + \boldsymbol{X_3} &\xrightarrow{k_i, \tau_i} 2\boldsymbol{X_2}, \\
\boldsymbol{X_2} &\xrightarrow{k_r, \tau_r} \boldsymbol{X_3}, \\
\boldsymbol{X_3} &\xrightarrow{k_s, \tau_s} \boldsymbol{X_1}
\end{aligned} \tag{55}$$

where $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$ denote the susceptible, infectious and recovered people, respectively. The first three rows represent the inflow and outflow of those people in the system, followed by the contamination reaction, the recovery reaction and the susceptibility reaction. We assume that only the three last reactions are affected by some delays. The stoichiometric matrix restricted to bimolecular reactions is given by $\begin{bmatrix} -1 & 1 & 0 \end{bmatrix}^T$ and hence Theorem 25 demands that $-v_1 + v_2 < 0$. We also have that $S_b = 0$ and

$$A = A_{\mathrm{df}} = \begin{bmatrix} -\gamma_1 & 0 & k_s \\ 0 & -\gamma_2 - k_r & 0 \\ 0 & k_r & -\gamma_3 - k_s \end{bmatrix} \tag{56}$$

where we have already removed the delayed unimolecular reactions and replaced them by their delay-free counterparts. A necessary condition for the feasibility of $v^T A_{\mathrm{df}} < 0$ is that the matrix A_{df} be Hurwitz stable. One can see that this matrix is structurally stable in the sense that it is Hurwitz stable for all positive values of its parameters. Indeed, one has that $\mathbb{1}^T A_{\mathrm{df}} < 0$ for all positive values of the parameters. To prove that it the network is also structurally ergodic in the presence of the delays, we need to verify $v^T A_{\mathrm{df}} < 0$ for some $v > 0$ such that $-v_1 + v_2 < 0$. Let $v = \begin{bmatrix} 1+\varepsilon & 1 & 1 \end{bmatrix}^T$, then we get that

$$v^T A_{\mathrm{df}} = \begin{bmatrix} -\gamma_1(1+\varepsilon) & -\gamma_2 & \varepsilon\, k_s - \gamma_3 \end{bmatrix}$$

and, hence, picking any $0 < \varepsilon < \gamma_3/k_s$ proves that the delayed reaction network is ergodic. Even stronger, the delayed reaction network is structurally ergodic.

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

For simplicity, let us consider here delayed reaction network $(\boldsymbol{X}, mathcalR, \tau)$ with zeroth- and first-order mass-action kinetics. Let us consider the associated augmented network $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup \mathcal{R}_d)$. The moment dynamics of the augmented network can be generically written as

$$\frac{d}{dt}\begin{bmatrix}\mathbb{E}[X(t)]\\ \mathbb{E}[D(t)]\end{bmatrix} = \begin{bmatrix}A & B\\ C & H^T\end{bmatrix}\begin{bmatrix}\mathbb{E}[X(t)]\\ \mathbb{E}[D(t)]\end{bmatrix} + \begin{bmatrix}b_0\\ b_d\end{bmatrix} \tag{57}$$

where the matrices are defined as in Section 5.2 together with

$$b_0 := S_x \left[\begin{array}{c}0\\ \hline b_x\end{array}\right] \text{ and } b_d = S_{\text{in},d}\left[\begin{array}{c}0\\ \hline b_{\text{in}}\end{array}\right]. \tag{58}$$

We then have the following result:

Proposition 27 *The dynamics of the moments first-order moments of the delayed network $(\boldsymbol{X}, \mathcal{R}, \tau)$ is described by*

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= A\mathbb{E}[X(t)] + \int_0^t S_{\text{out},x} f(s)\left[\begin{array}{c}W_{\text{in}}\mathbb{E}[X(t-s)]\\ \hline 0\end{array}\right] ds + b_0\\ &\quad + Be^{H^T t}\mathbb{E}[D(0)] + S_{\text{out},x}F(t)\left[\begin{array}{c}0\\ \hline b_{\text{in}}\end{array}\right]\end{aligned} \tag{59}$$

with $F(t) := \text{diag}_{i=1}^n(F_i(t))$ and $f(s) := \text{diag}_{i=1}^n(f_i(s))$ where $F_i(t) := 1-\alpha_i e^{H_i t}\mathbb{1}$ and $f_i(s) := -\alpha_i e^{H_i s}H_i\mathbb{1}$ are the cumulative distribution and the probability density function of the delay $\tau_i \sim \text{PH}(H_i, \alpha_i)$.

Proof : Solving for $\mathbb{E}[D(t)]$ in the moment system (57) yields

$$\mathbb{E}[D(t)] = e^{H^T t}\mathbb{E}[D(0)] + \int_0^t e^{H^T s}C\mathbb{E}[X(t-s)]ds + H^{-T}(I - e^{H^T t})b_d. \tag{60}$$

After substitution of the above expression in (57), we obtain

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X(t)] &= A\mathbb{E}[X(t)] + b_0 + Be^{H^T t}\mathbb{E}[D(0)] - H^{-T}(I - e^{H^T t})b_d\\ &\quad + B\int_0^t e^{H^T s}C\mathbb{E}[X(t-s)]ds.\end{aligned} \tag{61}$$

We have that

$$\begin{aligned}Be^{H^T s}C &= S_{\text{out},x}\operatorname*{diag}_{i=1}^{n}(-\mathbb{1}^T H_i^T e^{H_i^T s})\left[\begin{array}{c}\operatorname*{col}_{i=1}^{m}(\alpha_i^T q_i^T)\\ \hline 0\end{array}\right]\\ &= S_{\text{out},x}\left[\begin{array}{c}\operatorname*{col}_{i=1}^{m}(-\mathbb{1}^T H_i^T e^{H_i^T s}\alpha_i^T q_i^T)\\ \hline 0\end{array}\right]\\ &= S_{\text{out},x}\left[\begin{array}{c}\operatorname*{col}_{i=1}^{m}(f_i(s)q_i^T)\\ \hline 0\end{array}\right]\\ &= S_{\text{out},x}f(s)\left[\begin{array}{c}W_{\text{in}}\\ \hline 0\end{array}\right].\end{aligned} \tag{62}$$

where $f(s) = \mathrm{diag}_{i=1}^{n}(f_i(s))$ and $f_i(s) = -\alpha_i e^{H_i s} H_i \mathbb{1}$ is the probability density function of $\tau_i \sim \mathrm{PH}(H_i, \alpha_i)$. We also have that

$$\begin{aligned} -BH^{-T}(I - e^{H^T t})b_d &= S_{\mathrm{out},x} \operatorname*{diag}_{i=1}^{n}(\mathbb{1}^T H_i^T) H^{-T}(I - e^{H^T t}) \operatorname*{diag}_{i=1}^{n}(\alpha_i^T) \left[\begin{array}{c} 0 \\ \hline b_{\mathrm{in}} \end{array}\right] \\ &= S_{\mathrm{out},x} \operatorname*{diag}_{i=1}^{n}(\mathbb{1}^T (I - e^{H_i^T t})) \operatorname*{diag}_{i=1}^{n}(\alpha_i^T) \left[\begin{array}{c} 0 \\ \hline b_{\mathrm{in}} \end{array}\right] \\ &= S_{\mathrm{out},x} \operatorname*{diag}_{i=1}^{n}(\mathbb{1}^T (I - e^{H_i^T t}) \alpha_i^T) \left[\begin{array}{c} 0 \\ \hline b_{\mathrm{in}} \end{array}\right] \\ &= S_{\mathrm{out},x} F(t) \left[\begin{array}{c} 0 \\ \hline b_{\mathrm{in}} \end{array}\right] \end{aligned} \tag{63}$$

where $F(t) = \mathrm{diag}_{i=1}^{n}(F_i(t))$ and $F_i(t) = 1 - \alpha_i e^{H_i t} \mathbb{1}$ is the cumulative distribution function of $\tau_i \sim \mathrm{PH}(H_i, \alpha_i)$. The result follows. $\diamondsuit$

From the above expression, we can see that the first-order moment dynamics involves convolutions with kernels corresponding to the probability density functions of the delays. In this regard, the presence of phase-type distributed delays in the stochastic dynamics corresponds to filtering terms in 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

$$\int_0^t \delta(s - \bar{\tau}) x(t - s) ds = x(t - \bar{\tau}), t \geq \bar{\tau}. \tag{64}$$

Even if the result has been derived for unimolecular reaction networks, it readily generalizes to networks having more complex propensities at the expense of clarity.

6.2 Invariance of the stationary first-order moments

We have shown in the previous section that the dynamics of the first-order moments are affected by the presence of phase-type distributed delays. We have also proven that an augmented 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 for the augmented network coincides with that of the delay-free and the delayed network.

Proposition 28 *Let us consider here delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with zeroth- and first-order mass-action kinetics, and define its associated augmented network as $(\boldsymbol{X} \cup \boldsymbol{D}, \mathcal{R}^0 \cup \mathcal{R}_d)$. Assume that the Markov process associated with the augmented network is ergodic.*

Then, the stationary first-order moments of the augmented reaction network (and thus that of the delayed reaction network) coincide with the stationary first-order moments of the delay-free reaction network.

Proof : Since the Markov process associated with the augmented network is ergodic, then there exists a unique stationary distribution and the limit $\lim_{t \to \infty} \mathbb{E}[X(t)] = \mathbb{E}_\pi[X]$ holds re-

gardless of the initial conditions. Furthermore, it solves

$$\begin{aligned} 0 &= \lim_{t\to\infty}\left(A\mathbb{E}[X(t)] + \int_0^t S_{\text{out},x} f(s) \begin{bmatrix} W_{\text{in}}\mathbb{E}[X(t-s)] \\ \hline 0 \end{bmatrix} ds + b_0 \right. \\ &\quad \left. + Be^{H^T t}\mathbb{E}[D(0)] + S_{\text{out},x} F(s) \begin{bmatrix} 0 \\ \hline b_{\text{in}} \end{bmatrix}\right). \end{aligned} \tag{65}$$

This yields

$$\left(A + S_{\text{out},x}\begin{bmatrix} W_{\text{in}} \\ \hline 0 \end{bmatrix}\right)\mathbb{E}_\pi[X] + S_x \begin{bmatrix} b_x \\ \hline 0 \end{bmatrix} + S_{\text{out},x}\begin{bmatrix} 0 \\ \hline b_{\text{in}} \end{bmatrix} = 0 \tag{66}$$

where we have used the fact that $\int_0^\infty f(s)ds = I$ and $F(t) \to 0$ as $t \to \infty$. Finally, noting that the term in brackets is equal to A_{df}, the result follows. ♢

This result then demonstrates that phase-type distributed delay do not affect the stationary value of the first-order moments.

6.3 Variance is affected by the delays

In the previous section, we have shown that the stationary first-order moments of the network remain the same regardless the presence of the delays. However, the variance is influenced by the presence of the delay. Unfortunately, it seems difficult to prove that in the general setting. So, instead we will prove that in the case of a gene expression network with maturation delay. To this aim, let us consider the gene expression network

$$\begin{aligned} &\emptyset \xrightarrow{k_1} \boldsymbol{X_1} \xrightarrow{\gamma_1} \emptyset, \\ &\boldsymbol{X_1} \xrightarrow{k_2} \boldsymbol{X_1} + \boldsymbol{X_2}, \\ &\boldsymbol{X_2} \xrightarrow{\gamma_2} \emptyset \end{aligned} \tag{67}$$

and its delayed counterpart

$$\begin{aligned} &\emptyset \xrightarrow{k_1} \boldsymbol{X_1} \xrightarrow{\gamma_1} \emptyset, \\ &\boldsymbol{X_1} \xrightarrow{k_2} \boldsymbol{X_1} + \boldsymbol{D_1}, \\ &\boldsymbol{X_2} \xrightarrow{\gamma_2} \emptyset, \\ &\boldsymbol{D_1} \xrightarrow{\lambda} \boldsymbol{X_2} \end{aligned} \tag{68}$$

where $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ denote the mRNA and the protein species, respectively. The delay only consist of one phase here and is therefore exponentially distributed with rate λ. Calculations show that the protein variance is

$$V(X_2) = \mu\left(1 + \frac{k_2}{\gamma_1 + \gamma_2}\right) \tag{69}$$

in the delay-free case but is equal to

$$V_\lambda(X_2) = \mu\frac{\gamma_1\gamma_2(\gamma_1+\gamma_2) + \lambda(\gamma_1+\gamma_2)(\gamma_1+\gamma_2+k_2) + \lambda^2(\gamma_1+\gamma_2+k_2)}{(\gamma_1+\gamma_2)(\gamma_1+\lambda)(\gamma_2+\lambda)}. \tag{70}$$

Not surprisingly, we have that $V_\lambda(X_2) \to V(X_2)$ as $\lambda \to \infty$ since, in this case, there is no waiting time in the queue. To compare, we define the ratio

$$R(\lambda) = \frac{V_\lambda(X_2)}{V(X_2)} = \frac{\gamma_1\gamma_2(\gamma_1+\gamma_2)}{(\gamma_1+\lambda)(\gamma_2+\lambda)(\gamma_1+\gamma_2+k_2)} + \frac{\lambda(\gamma_1+\gamma_2)+\lambda^2}{(\gamma_1+\lambda)(\gamma_2+\lambda)}. \tag{71}$$

Calculations show that

$$R'(\lambda) = \frac{\gamma_1\gamma_2(\gamma_1 k_2 + \gamma_2 k_2 + 2k_2\lambda)}{(\gamma_1+\gamma_2+k_2)(\gamma_1+\lambda)^2(\gamma_2+\lambda)^2} \tag{72}$$

and, therefore, the function $V_\lambda(X_2)$ is strictly increasing and verifies

$$\mu < V_\lambda(X_2) < V(X_2) = \mu\left(1 + \frac{k_2}{\gamma_1+\gamma_2}\right) \tag{73}$$

for all $\lambda \in (0, \infty)$. This demonstrates that in the case of gene expression, the protein variance is influenced by the presence of the delay. Interestingly, one can see that the variance is smaller and that the queue has a filtering effect. Another interesting fact is the lower bound which imposes a lower bounds on how much the noise can be reduced and this lower bound corresponds to the variance of a Poisson process.

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 $(\boldsymbol{X}, \mathcal{R}, \tau)$ with first-order mass action kinetics using a delayed antithetic integral controller

$$\boldsymbol{Z_1} \xrightarrow{k,\tau(H_1,\alpha_1)} \boldsymbol{Z_1} + \boldsymbol{X_1}, \boldsymbol{X_\ell} \xrightarrow{\theta,\tau(H_n,\alpha_n)} \boldsymbol{X_\ell} + \boldsymbol{Z_2}, \emptyset \xrightarrow{\mu} \boldsymbol{Z_1}, \boldsymbol{Z_2} + \boldsymbol{Z_1} \xrightarrow{\eta} \emptyset. \tag{74}$$

Note that the annihilation reaction does not need to be delayed since it is a degradation reaction. Indeed, 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. Similarly, the reference reaction does not need to be delayed since from the point of view of the controller, only the output of the delay line will be visible and the actual content of the delay line does not matter.

The closed-loop network obtained from the interconnection of $(\boldsymbol{X}, \mathcal{R}, \tau)$ and the antithetic integral controller 74 is denoted by $(\boldsymbol{X} \cup \boldsymbol{Z}, \mathcal{R} \cup \mathcal{R}_{\text{AIC}}, \tau)$. The associated augmented and delay-free networks are denoted by $(\boldsymbol{X} \cup \boldsymbol{Z} \cup \boldsymbol{D}, \mathcal{R}^0 \cup \mathcal{R}_d \cup \mathcal{R}^0_{\text{AIC}} \cup \mathcal{R}_{\text{AIC},d})$ and $(\boldsymbol{X} \cup \boldsymbol{Z}, \mathcal{R} \cup \mathcal{R}_{\text{AIC}}, 0)$, respectively.

We assume that the actuated species $\boldsymbol{X_1}$ is produced after a stochastic time-varying delay modeled by the first delay-line whereas the sensing reaction is using the last delay

line. In this regard, we have that $q_1 = 0$, $q_n = \theta e_\ell$. For technical reasons, we assume that those delays are such that $\alpha_1, \alpha_n, H_1\mathbb{1}$ and $H_n\mathbb{1}$ have one and only one nonzero entry. This assumption is only very mildly restrictive. Based on these facts, we can define the following matrices

$$\begin{aligned} B &:= \left[-e_1\mathbb{1}^T H_1^T \;\middle|\; S_{\text{out},x}W_{\text{out}} \;\middle|\; 0\right] \\ C &= \left[\begin{array}{c} 0 \\ \hline \alpha_2^T q_2 \\ \vdots \\ \alpha_{n-1}^T q_{n-1} \\ \hline \alpha_n^T q_n \end{array}\right] \end{aligned} \tag{75}$$

where the first part corresponds the input delay, the middle part the delays intrinsic to the system and the last part of the measurement delay.

We have the following result:

Theorem 29 *Let us consider a delayed reaction network $(\boldsymbol{X}, \mathcal{R}, \tau)$ with only first-order mass-action kinetics and phase-type distributed delays. We assume that the associated delay-free network satisfies Assumption 3. Then, the following statements are equivalent:*

(a) The augmented closed-loop stochastic reaction network $(\boldsymbol{X} \cup \boldsymbol{Z} \cup \boldsymbol{D}, \mathcal{R}^0 \cup \mathcal{R}_d \cup \mathcal{R}^0_{\text{AIC}} \cup \mathcal{R}_{\text{AIC},d})$ is ergodic and we have that $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$.

(b) There exists some vectors $(v_x, v_\delta) \in \mathbb{R}^d_{>0} \times \mathbb{R}^{d_1+\ldots+d_n}$ and $(w_x, w_\delta) \in \mathbb{R}^d_{\geq 0} \times \mathbb{R}^{d_1+\ldots+d_n}$ such that the conditions

$$\begin{bmatrix} v_x \\ v_\delta \end{bmatrix}^T \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] < 0, \tag{76}$$

$$\begin{bmatrix} w_x \\ w_\delta \end{bmatrix}^T \left[\begin{array}{c|c} A & B \\ \hline C & H^T \end{array}\right] + \left[0 \;\middle|\; 0 \;\; \ldots \;\; -\mathbb{1}^T H_n^T\right] = 0 \tag{77}$$

and

$$\left[0 \;\middle|\; \alpha_1 \;\; 0 \;\; \ldots \;\; 0\right] w > 0 \tag{78}$$

hold.

(c) There exists some vectors $v_x \in \mathbb{R}^d_{>0}$ and $w_x \in \mathbb{R}^d_{\geq 0}$ satisfying $v_x^T e_1 > 0$, $w_x^T e_1 > 0$, $w_x^T e_\ell > 0$, such that the conditions

$$v^T A_{\text{df}} < 0 \text{ and } w^T A_{\text{df}} + e_\ell^T = 0 \tag{79}$$

hold.

(d) The delay-free closed-loop stochastic reaction network $(\boldsymbol{X} \cup \boldsymbol{Z}, \mathcal{R} \cup \mathcal{R}_{\text{AIC}}, 0)$ is ergodic and we have that $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$.

Proof : The conditions of the first statement are nothing else but the ergodicity and controllability conditions for the augmented reaction network. To prove the equivalence between the conditions, first note that we have that

$$\begin{bmatrix} v_x \\ v_\delta \end{bmatrix}^T \begin{bmatrix} A & B \\ C & H^T \end{bmatrix} < 0 \tag{80}$$

if and only if $v_x^T(A - BH^{-T}C) = v_x^T A_{\text{df}} < 0$. We also have that

$$\begin{bmatrix} w_x \\ w_\delta \end{bmatrix}^T \begin{bmatrix} A & B \\ C & H^T \end{bmatrix} + \begin{bmatrix} 0 & \mid & 0 & \dots & -\mathbb{1}^T H_n^T \end{bmatrix} = 0 \tag{81}$$

together with

$$\begin{bmatrix} w_x \\ w_\delta \end{bmatrix}^T \begin{bmatrix} 0 \\ \hline \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} > 0. \tag{82}$$

So, from $w_x^T B + w_\delta^T H^T = 0$, we get that

$$w_\delta^T = -(\begin{bmatrix} 0 & \mid & 0 & \dots & -\mathbb{1}^T H_n^T \end{bmatrix} + w_x^T B)H^{-T} \tag{83}$$

and hence

$$w_x^T(A - BH_n^{-T}C) + \begin{bmatrix} 0 & \dots & -\mathbb{1}^T H_n^T \end{bmatrix} H^{-T}C = 0 \tag{84}$$

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

$$\begin{aligned} w^T \begin{bmatrix} 0 \\ \hline \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} &= w_\delta^T \begin{bmatrix} \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} = -(\begin{bmatrix} 0 & \mid & 0 & \dots & -\mathbb{1}^T H_n^T \end{bmatrix} + w_x^T B)H^{-T} \begin{bmatrix} \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= w_x^T B H^{-T} \begin{bmatrix} \alpha_1^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= w_x^T e_1 \mathbb{1}^T H_1^T H_1^{-T} \alpha_1^T \\ &= w_x^T e_1 \end{aligned} \tag{85}$$

where we have used the fact that $\mathbb{1}^T \alpha_1^T = 1$. 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

$$
\begin{aligned}
\frac{d}{dt}\begin{bmatrix}\mathbb{E}[X(t)]\\ \hline \mathbb{E}[D(t)]\end{bmatrix} &= \left[\begin{array}{c|c} A & B \\ \hline C & H^T\end{array}\right]\begin{bmatrix}\mathbb{E}[X(t)]\\ \mathbb{E}[D(t)]\end{bmatrix} + \begin{bmatrix} 0 \\ \hline k\alpha_1^T \\ 0 \\ \vdots \\ 0\end{bmatrix}\mathbb{E}[Z_1(t)] \\
\frac{d}{dt}\mathbb{E}[Z_1(t)] &= \mu - \eta\mathbb{E}[Z_1(t)]\mathbb{E}[Z_2(t)] \\
\frac{d}{dt}\mathbb{E}[Z_2(t)] &= \left[0 \;\middle|\; 0 \;\; \dots \;\; 0 \;\; -\theta\mathbb{1}^T H_n^T\right]\begin{bmatrix}\mathbb{E}[X(t)]\\ \mathbb{E}[D(t)]\end{bmatrix} - \eta\mathbb{E}[Z_1(t)]\mathbb{E}[Z_2(t)].
\end{aligned}
\tag{86}
$$

The stationary solution is given by

$$
\begin{bmatrix}\mathbb{E}_\pi[X]\\ \mathbb{E}_\pi[D]\end{bmatrix} = -k\begin{bmatrix} A & B \\ C & H^T\end{bmatrix}^{-1}\begin{bmatrix} 0 \\ \hline \alpha_1^T k\mathbb{E}_\pi[Z_1] \\ 0 \\ \vdots \\ 0\end{bmatrix}
\tag{87}
$$

and

$$
\mu/\theta = \left[0 \;\middle|\; 0 \;\; \dots \;\; -\mathbb{1}^T H_n^T\right]\begin{bmatrix}\mathbb{E}_\pi[X]\\ \mathbb{E}_\pi[D]\end{bmatrix}
\tag{88}
$$

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

$$
\begin{bmatrix} A & B \\ C & H^T\end{bmatrix}^{-1} = \begin{bmatrix} (A-BH^{-T}C)^{-1} & (A-BH^{-T}C)^{-1}BH^{-T} \\ -H^{-T}C(A-BH^{-T}C)^{-1} & H^{-T}+H^{-T}C(A-BH^{-T}C)^{-1}BH^{-T}\end{bmatrix}
\tag{89}
$$

where

$$
\begin{aligned}
A - BH^{-T}C &= A_{\mathrm{df}} \\
H^{-T}C &= \begin{bmatrix} 0 \\ H_2^{-T}\alpha_2^T q_2^T \\ \vdots \\ H_n^{-T}\alpha_n^T q_n^T\end{bmatrix} \\
BH^{-T} &= = \left[-e_1\mathbb{1}^T H_1^T \quad S_{\mathrm{out},x}W_{\mathrm{out}} \quad 0\right].
\end{aligned}
\tag{90}
$$

Finally, we have that

$$
-\left[0 \;\middle|\; 0 \;\; \dots \;\; -\mathbb{1}^T H_n^T\right]\begin{bmatrix} A & B \\ C & H^T\end{bmatrix}^{-1}\begin{bmatrix} 0 \\ \hline \alpha_1^T k\mathbb{E}_\pi[Z_1] \\ 0 \\ \vdots \\ 0\end{bmatrix} = \mu/\theta
\tag{91}
$$

and, hence,

$$\begin{bmatrix}0 & \dots & 0 & \mathbb{1}^T H_n^T\end{bmatrix}\left(H^{-T}+H^{-T}C(A-BH^{-T}C)^{-1}BH^{-T}\right)\begin{bmatrix}\alpha_1^T k\mathbb{E}_\pi[Z_1]\\ 0\\ \vdots\\ 0\end{bmatrix}=\mu/\theta. \tag{92}$$

Using the explicit form for the matrices A, B, C and H yields $\mathbb{E}_\pi[Z_1] = \dfrac{\mu}{kg\theta}$ where $g := e_\ell^T(A-BH^{-T}C)^{-1}e_1$ is the static-gain of the network. 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

$$\mathbb{E}_\pi[X_\ell] = \mu/\theta. \tag{93}$$

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

As for the ergodicity analysis, we can see that the conditions also reduce to the condition for delay-free networks. In this regard, phase-type distributed delays are not harmful to the stability properties of the closed-loop network unlike in the deterministic setting.

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 [13,17,18]. 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 [36]. 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.

Another extension would be the study of the impact of delays on the variance and, in particular, the clarification of the conditions under which delays have a filtering effect on the noise. It may be tempting to say that this filtering mechanism relies on the randomness of the delay and would not be achieved by its deterministic counterpart. Elucidating this could help understand how cells filter their molecular noise and to design new synthetic filters topologies. Extrapolating a bit the result on gene expression, designing a filter that reduces noise may need to have small rates and, as a result, slow dynamics resulting in a slow response of the filter. In this regard, one can foresee the presence of some tradeoff between speed and quality of filtering.

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.

As a final comment, we would like to propose the following conjecture:

Conjecture 30 *Any delayed unimolecular stochastic reaction network with constant deter-*

ministic/stochastic delays or time-varying stochastic delays is ergodic if and only if its delay-free counterpart is.

We have proven under certain conditions that this conjecture is true when the delay is stochastic time-varying and phase-type distributed.

References

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

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

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

[4] S. Aoki, G. Lillacci, A. Gupta, A. Baumschlager, D. Schweingruber, and M. Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570:533–537, 2019.

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

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

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

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

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

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

[11] 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*, 102(42):14593–14598, 2005.

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

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

[14] C. Briat. Stability and performance analysis of linear positive systems with delays using input-output methods. *International Journal of Control*, 91(7):1669–1692, 2018.

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

[16] 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*, pages 92–95, Groningen, The Netherlands, 2014.

[17] 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*, pages 752–757, Las Vegas, USA, 2016.

[18] C. Briat and M. Khammash. Robust and structural ergodicity analysis of stochastic biomolecular networks involving synthetic antithetic integral controllers. In *20th IFAC World Congress*, pages 11405–11410, Toulouse, France, 2017.

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

[20] 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*, 449:103–123, 2018.

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

[22] S. Engblom. On the stability of stochastic jump kinetics. *ArXiv:1202.3892*, 2012.

[23] J. Falk, L. Bronstein, M. Hanst, B. Drossel, and H. Koeppl. Context in synthetic biology: Memory effects of environments with mono-molecular reactions. *Journal of Chemical Physics*, 150, 024106, 2019.

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

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

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

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

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

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

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

[31] C. C. Guet, A. Gupta, A. Henzinger, M. Mateescu, and A. Sezgin. Delayed continuous-time markov chains for genetic regulatory circuits. In P. Madhusudan and S. A. Seshia, editors, *Computer Aided Verification*, volume 7358 of *Lecture Notes in Computer Science*, pages 294–309. Springer Berlin Heidelberg, 2012.

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

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

[34] 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*, 147(154101):1–23, 2017.

[35] W. M. Haddad and V. Chellaboina. Stability theory for nonnegative and compartmental dynamical systems with time delay. *Systems & Control Letters*, 51(5):355–361, 2004.

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

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

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

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

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

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

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

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

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

[45] 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*, 11:20140108, 2014.

[46] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197:287–289, 1977.

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

[48] 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*, 102:068105, 2009.

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

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

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

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

[53] M. Rathinam. Moment growth bounds on continuous time Markov processes on non-negative integer lattices. *Quaterly of Applied Mathematics*, LXXIII(2):347–364, 2015.

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

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

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

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

[58] 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*, 456:516–519, 2008.

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

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