A Markovian Arrival Stream Approach to Stochastic Gene Expression in Cells

Brian Fralix* Mark Holmes† Andreas Löpker‡

December 22, 2022

Abstract

We analyse a generalisation of the stochastic gene expression model studied recently in Fromion et al. (*SIAM Journal of Applied Mathematics*, 2013) and Robert (*Probability Surveys*, 2019) that keeps track of the production of both mRNA and protein molecules, using techniques from the theory of point processes, as well as ideas from the theory of matrix-analytic methods. Here, both the activity of a gene and the creation of mRNA are modelled with an arbitrary Markovian Arrival Process governed by finitely many phases, and each mRNA molecule during its lifetime gives rise to protein molecules in accordance with a Poisson process. This modification is important, as Markovian Arrival Processes can be used to approximate many types of point processes on the nonnegative real line, meaning this framework allows us to further relax our assumptions on the overall process of transcription.

Keywords: infinite-server queues, Markov arrival process, matrix analytic methods, stochastic gene expression.

2020 MSC: 60G55, 60J28, 60K25, 92C40

1 Introduction

In this paper we are interested in analysing stochastic models of gene expression, which is the process through which genes within a cell produce both messenger RNA (mRNA) and protein molecules. Gene expression models date back to the 1970s with the works of Rigney and Schieve [46], Berg [6], and Rigney [44, 45], with later work occurring in the 1990s, see e.g. Ko [36] and Peccoud and Ycart [41].

Over the last 15-20 years, interest in creating/analysing suitable models of gene expression has increased considerably within the molecular biology community, with contributions being made by researchers from many different disciplines within not only the natural sciences, but also engineering and the mathematical sciences. Our primary objective in this paper is to show how certain techniques from applied probability can be used to analyze a broad class of gene expression models that includes many

*School of Mathematical and Statistical Sciences, Clemson University (bfralix@clemson.edu).

†School of Mathematics and Statistics, The University of Melbourne (holmes.m@unimelb.edu.au). MH's research is supported by Future Fellowship FT160100166 from the Australian Research Council.

‡Department of Computer Science and Mathematics, HTW Dresden, University of Applied Sciences (lopker@htw-dresden.de).

of the most well-studied models found in the literature. These techniques allow us to recover various known results while we also establish new and interesting results concerning the joint distributions and moments of the number of mRNA and protein molecules at each time $t \geq 0$.

In order to properly place more recent contributions to this research area in context, it will help to first introduce the three-stage model of gene expression found in the recent article of Robert [47]. This three-stage model is both a generalization of the classical stochastic model of gene expression (called the "telegraph model") presented in [41], and a special case of our model. In many mathematical analyses of protein production in cells, the production process of both mRNA and protein molecules by a fixed gene (a fixed segment of DNA) in a cell is considered. This process is often described via the evolution through time of a random vector $(I(t), M(t), P(t))$, where $I(t) \in \{0, 1\}$ describes the state of the gene at time t, and $M(t)$ and $P(t)$ denote the number of mRNA and protein molecules, respectively, present in the cell at time t. When $I(t) = 1$ the gene is said to be *active* at time t; otherwise (when $I(t) = 0$) the gene is *inactive* at this time. When the gene is active, an RNA polymerase may bind to it and start to copy the gene via sequential production of nucleotides: once this elongation process is complete, a new messenger RNA (mRNA) molecule is formed (so $M(t)$ increases by 1 if this process completes at time t). Molecular biologists refer to the overall process of creating new mRNA molecules from segments of DNA as *transcription*. mRNA molecules currently in the process of being transcripted are often referred to as *nascent* mRNA molecules. Once the elongation phase of a nascent mRNA molecule is complete, it becomes a *mature* mRNA molecule. In many stochastic models of gene expression, it is assumed that the elongation time associated with the production of each mRNA molecule is negligible, meaning all such elongation times are set equal to zero, but models have been created recently that allow for such elongation times to be both nonzero and random. In fact, it has been argued that tracking nascent mRNA may be more important than tracking mature mRNA molecules, see e.g. Choubey et al. [13], Choubey [12] and Filatova et al. [15].

Once a mature mRNA molecule has been created, a ribosome may in turn bind with it and move along its chain of nucleotides to build a protein (so $P(t)$ increases by 1) via a chain of amino acids: this process is referred to as *translation*. Again, in many models of gene expression it is assumed that the amount of time it takes a ribosome to produce a protein molecule, once it manages to bind to the mRNA molecule is negligible, but some recent models have incorporated random production times.

Both proteins and mRNA molecules are assumed to have finite, random lifetimes (at the end of such a lifetime $M(t)$ or $P(t)$ decreases by 1). The special case of Robert's three-stage model where lifetimes are all exponential, and all elongation times are zero is referred to as the "classical three-stage model" see e.g. Paulsson [40]. Due to the relatively low numbers of both mRNA and protein molecules within a typical cell, the process of gene expression is often modeled with a continuous-time, discrete-state stochastic processes [40]. We refer readers seeking a textbook-level discussion on stochastic gene expression to Chapter 6 of Bressloff [9]: another text that could be of interest to such readers is Anderson and Kurtz [2].

1.1 Literature Review

Stochastic gene expression is a young but extremely active area of research, and a host of models have been proposed in the literature to attempt to capture various facets of transcription and translation. In the interest of eventually comparing and contrasting our results with what others have done in this area, we provide a brief summary of the types of phenomena other gene expression models have attempted

to capture. More precisely we will discuss a substantial subset of the vast literature that we think is most relevant to the present work.

In [16, 47], the authors advocate that the three-stage model of stochastic gene expression can be fruitfully analysed with the theory of *Poisson random measures*. Of particular relevance to our work is their use of ideas that are heavily related to the theory of infinite-server queueing systems fed by Poisson arrival processes. One major advantage of this approach is that it facilitates the removal of the exponential assumption in various places within the model. From this viewpoint one can derive many interesting quantities (e.g. means and variances of mRNA and protein counts) even when all elongation times, mRNA lifetimes and protein lifetimes are allowed to be generally distributed.

Our primary objective is to complement the approach found in [16, 47] by showing how other established results/tools from the applied probability community can be used to analyse, rather thoroughly, various generalizations of the telegraph model. These generalisations allow for the gene to vary between activity and inactivity in accordance with an environment process (called a Markovian Arrival Process, or MAP) that is much more general than a Markov-modulated Poisson process governed by a two-state continuous-time Markov chain (CTMC). Our methods will involve a combination of both point process theory and ideas from a subject often referred to in applied probability as *matrix-analytic methods*. These methods were originally designed to find computable quantities that describe the behavior of queueing systems where interarrival times and/or service times are not exponentially distributed. Many types of stochastic systems having parameters that vary with time due to the behavior of some background environmental process can be studied with methods from the matrix-analytic framework, meaning this theory could potentially provide the molecular biology community with useful tools for deriving results that go beyond the results discussed here.

In the next two sections we present existing studies that fall within our general framework and outside our general framework respectively.

1.1.1 Studies that fall in our framework

The telegraph model of [41] appears to have been inspired by the ideas given in [36], and it is arguably the most well-known Markovian model of gene expression. In the telegraph model, the stochastic process $\{I(t); t \geq 0\}$ is assumed to be a two-state continuous-time Markov chain (CTMC), the lifetime of each mRNA molecule is assumed to be exponentially distributed and independent of everything else, and the elongation time associated with each mRNA molecule is assumed to be negligible. Protein molecules are not included in this model. The main results of [41] include (a) a study of the time-dependent distribution of the number of mRNA molecules when it is further assumed that each mRNA molecule has an infinite lifetime; (b) a derivation of the time-dependent mean and variance when each mRNA molecule has an exponentially distributed lifetime, and (c) a derivation of the equilibrium distribution of the number of mRNA molecules when each mRNA molecule has an exponentially distributed lifetime. Later, in Iyer-Biswas et al. [25], the authors studied the time-dependent distribution of the number of mRNA molecules in the telegraph model.

There are many ways in which the telegraph model can be generalised in order to make it more realistic. One direction involves extending the phase space of the gene from two states to an arbitrarily finite number of states, as a way of loosening the assumption that whenever the gene is active/inactive, it stays in that state for an exponentially distributed amount of time before becoming inactive/active. In the work of Zhou and Zhang [50], the authors assume the states of the gene vary with time in accordance

to a finite-state CTMC whose transition diagram is that of a circular Markov chain, see e.g. Adan and Resing [1]. The work of Herbach [22] further generalizes the model found in [50], by allowing the states of the gene to vary in accordance to a finite-state CTMC with an arbitrary transition structure. In both [50, 22], while the gene is in an active state, mRNA molecules are created in accordance to a Poisson process with some rate and the elongation phase is assumed to be negligible, but mRNA molecules are never created at any instants when the gene makes a phase transition. Another recent paper related to [50, 22] is Jia and Li [33], where time-dependent distributions are studied.

In Ham et al. [19] the authors present a generalization of the telegraph model that admits a tractable stationary (and time-dependent) distribution. In Cao et al. [10] the authors present another generalization of the telegraph model that keeps track of both nascent and mature mRNA molecules, where the gene fluctuates between three states: a nonpermissive state, and two different permissive states.

None of the papers cited in the previous three paragraphs model both mRNA molecules and protein molecules simultaneously. In the work of Shahrezaei and Swain [48] the authors study the joint distribution of $M(t)$ and $P(t)$ while further assuming the gene is always active: their analysis involves setting up the Kolmogorov forward equations associated with the continuous-time Markov chain that captures the dynamics of this model, then solving the equations with the method of characteristics under the further assumption that the lifetime of a protein molecule is typically much longer than the lifetime of an mRNA molecule (this imprecise statement is made precise in their paper). This method results in an expression for the joint generating function of $M(t)$ and $P(t)$ that is in terms of a hypergeometric function. A similar analysis is also used in [48] to study the joint distribution of $I(t), M(t)$, and $P(t)$, again under the assumption that the lifetime of a protein molecule is typically much longer than the lifetime of an mRNA molecule. A few years later, in the work of Bokes et al. [7] the authors found the joint generating function of the mRNA and protein levels for the case where the gene is always active, without having to assume mRNA lifetimes are much smaller than protein lifetimes.

Another stream of literature is centered around the idea of introducing nonzero elongation times within the transcription process and/or the translation process, without keeping track of the number of nascent mRNA molecules. Recent work on stochastic gene expression has focused on relaxing assumptions typically made about mRNA lifetimes, protein lifetimes, and elongation times. For example, the model that is the primary focus of [47] (which is a generalisation of that from Fromion et al. [16]) allows for elongation times and lifetimes with general distributions, with families of Poisson processes corresponding to possible binding times. Bindings occur at these possible binding times only if they occur while the gene is active (for the first kind of binding above) and during the lifetime of the mRNA molecule (for the second kind of binding above) respectively. In the work of Jansen and Pfaffelhuber [26] which predates [47], the authors model gene expression via random time changes of independent Poisson processes, from which they derive the mean and variance of both mRNA and protein levels, when the gene alternates between being active and inactive in accordance with a two-state CTMC, and when the lifetimes of each mRNA and each protein are exponentially distributed. It is notable that in their analysis, generally distributed elongation times are permitted.

Recently, experiments have been performed that suggest an improved understanding of gene expression can be gained by tracking nascent mRNA molecules. Early papers on this topic that present mathematical models while having this goal in mind include Choubey et al. [13] and Choubey [12]. In Filatova et al. [15], the authors essentially model elongation times as a sum of L independent, exponentially distributed random variables, so that each such nascent mRNA molecule must progress through

L exponential phases before becoming a mature mRNA molecule. There they consider a Markovian model that keeps track of the number of nascent mRNA models, and they derive various steady-state moments of these random variables. The work of Szavits-Nossan and Grima [49] further builds on [15] by considering a mean-field model in order to approximate various distributions associated with this model. In the recent work of De Gunst et al. [18] the authors study a generalisation of the telegraph model where mRNA molecules are conceived in accordance with a Poisson process while the gene is in an active state, but that as soon as an mRNA molecule is conceived, further conception is blocked until the newly-conceived molecule becomes a mature mRNA molecule. In [18], the elongation time of each mRNA molecule is assumed to be hypoexponentially distributed (meaning its distribution is simply the distribution of the sum of a finite number of independent, exponentially distributed random variables).

The results we describe in Section 4 generalise the results found in [22], as our framework does not require us to assume mRNA molecules are created by a Markov-modulated Poisson process. Indeed, MAPs can theoretically be used to approximate many types of point processes on the nonnegative real line–see Asmussen and Koole [3]–so we feel as though this could be an important step towards analyzing even more general models of stochastic gene expression.

It is interesting to note that other research activities in stochastic gene expression have involved ideas and techniques from the theory of matrix-analytic methods. Horowitz and Kulkarni [24] have recently used Batch Markovian Arrival Processes (BMAP) to study bursts of mRNA molecules. They argue that in their setting it is reasonable to model protein production with a BMAP (where proteins are created in batches in accordance with a MAP) for the case where mRNA lifetimes are significantly shorter than protein lifetimes. The authors of [24] focus more on studying rare events associated with their model, but unlike their model, ours is a straight generalization of the classical three-stage model of gene expression, where the activity/inactivity of the gene is modelled using a MAP, and all elongation times and lifetimes are generally distributed. Finally, the analysis featured in [18] involves the use of recently established results from the theory of level-dependent quasi-birth-death processes, which is significantly different from our overall approach. We will illustrate how our results shed additional light on the behavior of this model in Section 4.3.

1.1.2 Phenomena that our models do not explicitly address

Another subarea of research in this field involves constructing reasonable models that attempt to capture naturally occurring feedback mechanisms that appear in the process of gene expression. For example, each instant at which the gene becomes active is the result of an existing protein molecule binding with the gene, so it would be reasonable to modify the model so that the rate at which the gene switches from being inactive to being active depends on the number of protein molecules present in the cell. Relevant literature addressing this phenomenon include Hornos et al. [23], Kumar et al. [37], Jia and Grima [31], Jia et al. [29], and Jia et al. [28].

Another important phenomenon to consider is cell division. During the lifetime of a cell, a strand of DNA will replicate, and soon after the cell will divide into two cells. During the process of cell division, each molecule will either stay with the current cell, or move with the newly created cell (the daughter cell). In the work of Beentjes et al. [5], the authors propose a series of stochastic models that incorporate cell division, where the most general model studied in [5] that allows for a random cell cycle models the cell cycle as being hypoexponentially distributed. This work was followed by

Perez-Carrasco et al. [42] where the authors further build on the results from [5] by also incorporating DNA replication in the model. Other previous papers featuring a study of cell division include Jędrak et al. [27] and Jia and Grima [32].

Recently, there have been studies that attempt to feature two or more of the phenomena mentioned above in models of stochastic gene expression, see e.g. Cao and Grima [11] as well as Dessalles et al. [14]. In Ham et al. [20], various underlying parameters of the telegraph model are also allowed to vary randomly with time. Another generalization of the telegraph model has been introduced by Jia [30], in order to model the so-called *dropout effect*, which occurs when the mRNA level in a cell is measured to be zero, even if mRNA molecules are present in the cell.

Finally we note that (as is the case for all of the literature described in Section 1.1.1) our model does not allow interruptions during elongation periods. For example, if the gene switches to inactive during the elongation period of an mRNA molecule, the production of that molecule is not interrupted.

1.2 Organisation

This paper is organised as follows. In Section 2, we define the MAP, and we state our main results that describe the joint distribution of the number of nascent mRNA molecules, mature mRNA molecules, and protein molecules at each time $t \geq 0$. In Section 3 we prove our main results, and we conclude the paper in Section 4 by showing how more can be said about the distribution of mRNA molecules at each time $t \geq 0$ when the lifetime of each mRNA molecule is exponentially distributed, and each elongation time is negligible. We also provide many new results associated with the model studied in [18], and we verify our results through numerical work and simulation.

2 The model and main results

In this paper we analyse an important extension of the three-stage model of gene expression from [47]. Here we assume RNA polymerases bind with an active gene at random points of a so-called Markovian Arrival Process (MAP)—see below—having finite phase set S, meaning the transcription process of the gene is governed by the MAP. Associated with each created mRNA molecule is its elongation time (the time it takes for the mRNA molecule to be created from the RNA polymerase that binds with the gene) and its lifetime: the elongation time is generally distributed with CDF E_1, and the lifetime is generally distributed with CDF L_1. Furthermore, during the lifetime of an mRNA molecule, ribosomes bind to it in accordance with a homogeneous Poisson process having rate k_2. Each created protein has associated with it both an elongation time (the time it takes a ribosome that binds with the mRNA molecule to produce the protein molecule) and a lifetime: the elongation time of a protein is generally distributed with CDF E_2, while the lifetime of the protein is generally distributed with CDF L_2. Throughout we assume that all elongation times and lifetimes are independent of each other, as well as the MAP that governs when transcription is initiated.

Our main results are formulae describing the (joint) distribution of $M_{\text{nas}}(t)$, $M(t)$, and $P(t)$, and their moments, where $M_{\text{nas}}(t)$ denotes the number of nascent mRNA molecules present at time t. To be more precise, for $x \in S$ we will let $\mathbb{P}_x$ denote a probability measure under which this process evolves from initial conditions $M_{\text{nas}}(0) = M(0) = P(0) = 0$ and $I(0) = x$. Then for each $t \in \mathbb{R}_+ := [0, \infty)$,

each $m_1, m_2, n \in \mathbb{Z}_+ := \{0, 1, 2, \dots\}$, we define the matrix $\mathbf{J}(t, m_1, m_2, n)$ as

$$\mathbf{J}(t, m_1, m_2, n) := \left[\mathbb{P}_x(M_{\text{nas}}(t) = m_1, M(t) = m_2, P(t) = n, I(t) = y)\right]_{x,y\in S}. \tag{1}$$

For fixed x this gives the joint probability mass function (PMF) of $(M_{\text{nas}}(t), M(t), P(t), I(t))$. Letting $\mathbf{I}$ denote the $|S| \times |S|$ identity matrix, and $\mathbf{0}$ the $|S| \times |S|$ zero matrix, note that for each $m_1, m_2, n \in \mathbb{Z}_+$,

$$\mathbf{J}(0, m_1, m_2, n) = \begin{cases} \mathbf{I}, & m_1 = m_2 = n = 0; \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

A more compact description of the joint distribution of $(M_{\text{nas}}(t), M(t), P(t), I(t))$ can be made through the use of probability generating functions (PGFs). For each $t \in \mathbb{R}_+$, each $z_1, z_2, z_3 \in \mathbb{D} := \{z \in \mathbb{C} : |z| \leq 1\}$, and each integer $n \geq 0$ we define the matrix

$$\hat{\mathbf{J}}_n(t, z_1, z_2, z_3) := \left[\mathbb{E}_x[z_1^{M_{\text{nas}}(t)} z_2^{M(t)} z_3^{P(t)} \mathbb{1}_{\{I(t)=y, N(t)\leq n\}}]\right]_{x,y\in S}, \tag{2}$$

where $\mathbb{E}_x$ denotes expectation with respect to $\mathbb{P}_x$, and $\mathbb{1}_A$ denotes the indicator of A (i.e. $\mathbb{1}_A = 1$ if the event A occurs, and $\mathbb{1}_A = 0$ otherwise). The Dominated Convergence Theorem shows that as $n \to \infty$, $\hat{\mathbf{J}}_n$ converges pointwise to $\hat{\mathbf{J}}$, where

$$\hat{\mathbf{J}}(t, z_1, z_2, z_3) := \left[\mathbb{E}_x[z_1^{M_{\text{nas}}(t)} z_2^{M(t)} z_3^{P(t)} \mathbb{1}_{\{I(t)=y\}}]\right]_{x,y\in S}. \tag{3}$$

We can also study the moments (and cross-moments) of $M_{\text{nas}}(t)$, $M(t)$, and $P(t)$ directly via the matrices

$$\mathbf{C}_{m_1,m_2,n}(t) := \left[\mathbb{E}_x[M_{\text{nas}}(t)^{m_1} M(t)^{m_2} P(t)^n \mathbb{1}_{\{I(t)=y\}}]\right]_{x,y\in S}, \quad m_1, m_2, n \in \mathbb{Z}_+. \tag{4}$$

Our main results include formulae for $\mathbf{J}$, $\hat{\mathbf{J}}$, and $\mathbf{C}$ in Theorems 2.4, 2.5, and 2.6, respectively. One could in principle use either Theorem 2.4 or Theorem 2.5 to prove the others, but we give a direct proof (utilising point process theory) in each case. Section 3 contains the proofs of our main results from Section 2, and in Section 4, we use Theorem 2.5 to say more about the joint distribution of $I(t)$ and $M(t)$, when L_1 is the CDF of an exponentially distributed random variable, and all elongation times are equal to zero.

2.1 Markovian Arrival Processes

A MAP is a continuous-time Markov chain $\{(N(t), I(t)); t \geq 0\}$ having state space $\mathbb{Z}_+ \times S$ and transition rate matrix $\mathbf{Q}$, with S being a finite set. We assume without loss of generality that $S := \{0, 1 \dots, p\}$ for some integer $p \geq 0$, so the state space can be ordered lexicographically, and we order the states in this way when we construct the rows and columns of $\mathbf{Q}$. Here $N(t) \in \mathbb{Z}_+$ is a counting process (in our gene expression model $N(t)$ represents the number of mRNA polymerase bindings that occur in the interval $(0, t]$) and $I(t) \in S$ denotes the phase that the system is in at time t, which in our model corresponds to the state of the gene at time t (e.g. in the three-stage model of gene expression analysed in [47] $S = \{0, 1\}$, where $I(t) = 0$ means that the gene is inactive). The transition rates of such a process are governed by a pair of $|S| \times |S|$ matrices, $\mathbf{K}_0$ and $\mathbf{K}_1$. The off-diagonal entries of $\mathbf{K}_0$ (are non-negative and) correspond to transition rates associated with phase transition instants that are not counted by the counting process. Similarly, the off-diagonal entries of $\mathbf{K}_1$ correspond to transition rates

associated with phase transitions that are counted by the counting process, and the diagonal entries of $\mathbf{K}_1$ correspond to instants at which the counting process increases by one unit, without there being a change in the phase process. In queueing models, transition instants corresponding to rates from $\mathbf{K}_1$ often represent arrival times of customers to a queueing system, which is where the name *Markovian Arrival Process* comes from. Finally, the diagonal entries of $\mathbf{K}_0$ are then chosen so that each row sum of $\mathbf{Q}$ is zero, and due to $\mathbb{Z}_+ \times S$ being lexicographically ordered, $\mathbf{Q}$ can be expressed as follows:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{K}_0 & \mathbf{K}_1 & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{K}_0 & \mathbf{K}_1 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_0 & \mathbf{K}_1 & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_0 & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

We refer to $\{(N(t), I(t)); t \geq 0\}$ as the $(\mathbf{K}_0, \mathbf{K}_1)$-MAP. MAPs play a prominent role in what is known in applied probability as the theory of matrix-analytic methods: a textbook-level treatment of these topics is given in e.g. Latouche and Ramaswami [38] and He [21]. One way to construct a MAP is to do so by thinking of it as being governed by a finite collection of independent, homogeneous Poisson processes, where each Poisson process is associated with a particular element from $\mathbf{Q}$ (i.e. from $\mathbf{K}_0$ or $\mathbf{K}_1$), where

$$\mathbf{K}_0 := [k_0(z,w)]_{z,w\in S}, \qquad \mathbf{K}_1 := [k_1(z,w)]_{z,w\in S}.$$

Each off-diagonal (i.e. $k \neq z$) element $k_0(z,w)$ has associated with it a homogeneous Poisson process $\{A^{(0)}_{z,w}(t); t \geq 0\}$ with rate $k_0(z,w)$, and each element $k_1(z,w)$ has associated with it a homogeneous Poisson process $\{A^{(1)}_{z,w}(t); t \geq 0\}$ with rate $k_1(z,w)$. All of these Poisson processes are independent of each other, and they provide all of the randomness needed to construct $\{(N(t), I(t)); t \geq 0\}$ (from a given initial state). The evolution of $I(\cdot)$ is determined by the processes $\mathcal{I} = \{A^{(i)}_{z,w} : i \in \{0,1\}, z, w \in S, z \neq w\}$: when the current state is $I(t) = \ell$ the next transition of $I(\cdot)$ occurs at the next firing time among the processes $A^{(\cdot)}_{\ell,\cdot} \in \mathcal{I}$ (i.e. those with $z = \ell$), and the new state is the $w \neq \ell$ whose Poisson process fired at this time. In particular, $\{I(t); t \geq 0\}$ is a CTMC having generator matrix $\mathbf{K}_0+\mathbf{K}_1$. Readers interested in a more rigorous construction that shows how a collection of independent, homogeneous Poisson processes can be used to govern CTMC are referred to Chapter 9 of Brémaud [8]. Note that it is possible to modify this model so that a random number of mRNA molecules can be conceived simultaneously. This requires replacing the MAP with a so-called Batch Markovian Arrival Process (BMAP), having a generator matrix that can be expressed in block-partitioned form in terms of not just $\mathbf{K}_0$ and $\mathbf{K}_1$, but other matrices $\mathbf{K}_2, \mathbf{K}_3, \mathbf{K}_4, \ldots$. All of our main results can be modified to this more general setting, but we will refrain from doing this in the interest of preserving readability.

The $N(t)$ process can be constructed similarly from the full set of Poisson processes (or from the $I(t)$ process and the $A^{(1)}$ Poisson processes). For each $t \geq 0$,

$$N(t) := \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{I(s-)=z\}} A^{(1)}_{z,w}(ds)$$

so that $N(t)$ counts the number of firings (up to time t) of $A^{(1)}_{z,\cdot}$ processes that occur while I is in the corresponding state z, summed over $z \in S$. Readers should interpret each integral encountered in this paper as a Lebesgue-Stieltjes integral. In particular, we use the interval notation in the integral because the function of integration (i.e. $A^{(1)}_{z,w}$ in the expression above) may have a jump at an endpoint of I.

Example 2.1. The telegraph model of Peccoud and Ycart [41] keeps track of mRNA created by a single gene that alternates between being active and inactive in accordance with a two-state CTMC with state space $S = \{0, 1\}$. While the gene is active, mRNA molecules are created in accordance with a Poisson process, and no mRNA molecules are created while the gene is inactive. We can express this process in MAP language (while also using the notation found in [19]) as

$$\mathbf{K}_0 = \begin{pmatrix} -\lambda & \lambda \\ \mu & -(K_A + \mu) \end{pmatrix}, \qquad \mathbf{K}_1 = \begin{pmatrix} 0 & 0 \\ 0 & K_A \end{pmatrix}.$$

where K_A is the rate at which RNA polymerase bindings occur with the (active) gene, λ is the rate at which the gene (when inactive) becomes active, and μ is the rate at which the gene (when active) becomes inactive. Readers should note that in the telegraph model of [41], each mRNA molecule has a negligible elongation time and an exponentially distributed lifetime with rate δ, independently of everything else, but in our general setting elongation times and lifetime are modelled separately from the MAP governing gene activity and mRNA conception.

Observe too that in this example $(I(t))_{t\geq 0}$ is a 2-state CTMC with generator $\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$, and

$$N(t) = \int_{(0,t]} \mathbb{1}_{\{I(s-)=1\}} A_{1,1}^{(1)}(ds).$$

Example 2.2. The model analyzed in [10] corresponds to $S = \{0, 1, 2\}$ and

$$\mathbf{K}_0 = \begin{pmatrix} -a_1 & a_1 & 0 \\ a_0 & -(a_0 + a_2) & a_2 \\ a_0 & 0 & -(a_0 + \rho) \end{pmatrix}, \qquad \mathbf{K}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \rho & 0 \end{pmatrix}.$$

In the notation of [10] $a_1 = \sigma_u$, $a_0 = \sigma_b$ and $a_2 = \lambda$. In this model the gene alternates between three possible states. Whenever the gene makes a transition from state 2 to state 1, an mRNA molecule is conceived. We will analyze this model further in Section 4.

To illustrate the flexibility of this class of processes, consider the following example.

Example 2.3. Let $S = \{0, 1, 2\}$ and

$$\mathbf{K}_0 = \begin{pmatrix} -a_1 & a_1 & 0 \\ \lambda_- & -(\rho_1 + \lambda_- + \lambda_+) & \lambda_+ \\ b_0 & b_1 & -(b_0 + b_1 + \rho_2 + \rho_{2,1}) \end{pmatrix}, \qquad \mathbf{K}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \rho_1 & 0 \\ 0 & \rho_{2,1} & \rho_2 \end{pmatrix}.$$

An interpretation of this MAP in terms of a mRNA production model could be as follows: When the gene is inactive there are no RNA polymerase bindings. There are two "active" phases. The (inactive) gene becomes active-1 at rate a_1, and when active-1 the gene becomes inactive at rate λ_- and active-2 with rate λ_+ (for each such transition no RNA polymerase bindings occur). Such bindings do occur at rate ρ_1 while the gene is active-1 and at rate $\rho_2+\rho_{2,1}$ while the gene is active-2. The rate ρ_2 corresponds to bindings while the gene stays as active-2. The rate $\rho_{2,1}$ corresponds to transitions where a binding occurs with the active-2 gene and the gene immediately becomes active-1. While active-2, the gene may also become active-1 (at rate b_1) or inactive (at rate b_0) via transitions that do not produce an RNA molecule.

2.2 Main Results

The main results of this paper will address the following three-stage model of stochastic gene expression, which behaves in accordance with the following rules:

(A1) the transcription process within a cell is assumed to be governed by a MAP $\{(N(t), I(t)); t \geq 0\}$ having finite phase set S and matrices $\mathbf{K}_0$ and $\mathbf{K}_1$, where each jump point of N corresponds to an instant where an RNA polymerase binds with the (active) gene to begin producing an mRNA molecule,

(A2) for each such binding, the amount of time it takes to create an mRNA molecule is generally distributed with CDF E_1, and independent of everything else (let $(D^{(i)})_{i\in\mathbb{N}}$ denote these (i.i.d.$\sim E_1$) elongation times),

(A3) the lifetime of each mRNA is assumed to be generally distributed with CDF L_1, and independent of everything else (let $(B^{(i)})_{i\in\mathbb{N}}$ denote these (i.i.d.$\sim L_1$) lifetimes),

(A4) while an mRNA exists, it initiates the creation of proteins in accordance with a homogeneous Poisson process having rate $k_2 > 0$, independent of everything else,

(A5) when an mRNA initiates the creation of a protein, it takes a generally distributed amount of time having CDF E_2 to actually create the protein, with this creation time being independent of everything else,

(A6) each protein exists in the cell for a generally distributed amount of time, having CDF L_2, independent of everything else.

We also assume throughout that the cdfs L_1, L_2, E_1, and E_2 are all proper, and correspond to nonnegative random variables: saying e.g. L_1 satisfies this criteria means $\lim_{x\uparrow 0} L_1(x) = 0$ and $\lim_{x\to\infty} L_1(x) = 1$. In fact, in order to guarantee that the random vector $(I(t), M_{\text{nas}}(t), M(t), P(t))$ converges in distribution as $t \to \infty$ to a non-degenerate limit we will also need to assume the random variables associated with E_1, L_1, E_2, and L_2 have finite means.

It is important to point out that the MAP only determines the instants at which a RNA polymerase binds with the (active) gene. For instance, if the gene switches from being active to being inactive while an mRNA is currently going through its elongation phase, that elongation phase is not affected in any way, nor is the lifetime following that phase. This may or may not be realistic, but this is how elongation phases behave in both [26] and [47]. The three-stage model of gene expression from [40, 48] can be described exactly within this framework: simply choose $\mathbf{K}_0$ and $\mathbf{K}_1$ to be the matrices given in our description of the telegraph model from [41], choose E_1 and E_2 to satisfy $E_1(0-) = E_2(0-) = 0$ and $E_1(0) = E_2(0) = 1$ (i.e. elongation times are 0), and let L_1 and L_2 be the CDFs of exponentially distributed random variables, having rates γ_1 and γ_2, respectively.

Before we begin stating our mathematical results, we first need to introduce some additional notation, much of which is standard. Given a proper CDF F defined on $\mathbb{R}$, we let $\overline{F}$ denote its corresponding tail function, meaning $\overline{F}(x) := 1 - F(x)$ for each $x \in \mathbb{R}$. Recall also that the CDF of the sum of two independent nonnegative random variables with CDFs F and G is given by the convolution $F \star G$, which is defined as

$$(F \star G)(t) := \int_{[0,t]} G(t-s)dF(s), \quad t \geq 0.$$

As we proceed through this paper, we will require an understanding of the distribution of the number of proteins present in the system at time $s+t$ that were produced by an RNA polymerase binding event that occurred at time s. This distribution depends on t but not s, so without loss of generality we consider the case $s=0$ in what follows. Associated with an RNA polymerase that binds with the (active) gene at time zero is the (elongation) time $D_0 \sim E_1$ that it takes to create an mRNA molecule, and $B_0 \sim L_1$, the lifetime of that mRNA molecule. Finally, let $P(\{0\},t)$ denote the number of proteins created by that mRNA molecule that are present in the cell at time t. We assume that this "bind at time 0" process is defined on some space with probability measure $\mathcal{P}_*$ and associated expectation $\mathcal{E}_*$.

Recall the definition (1), and note that $\mathbf{J}(0,m_1,m_2,n) = \mathbf{I}\mathbb{1}_{\{m_1=m_2=n=0\}}$, where $\mathbf{I}$ denotes the identity matrix. Our first main result shows that the matrices $\mathbf{J}(t,m_1,m_2,n)$ satisfy a relatively simple recursion. To state this as briefly as possible it is convenient to define $\mathbf{J}(t,-1,m_2,n) = \mathbf{0} = \mathbf{J}(t,m_1,-1,n)$ for every $m_1,m_2,n \in \mathbb{Z}_+$ and $t \geq 0$.

Theorem 2.4. *Assume* (A1)-(A6). *For Lebesgue almost every $t>0$, the following holds*:
For each $m_1,m_2,n \in \mathbb{Z}_+$,

$$\begin{aligned}\frac{\partial}{\partial t}\mathbf{J}(t,m_1,m_2,n) &= \mathbf{K}_0\mathbf{J}(t,m_1,m_2,n) + \overline{E}_1(t)\mathbf{K}_1\mathbf{J}(t,m_1-1,m_2,n)\\ &\quad + \sum_{k=0}^{n}\mathcal{P}_*(D_0\leq t, D_0+B_0>t, P(\{0\},t)=n-k)\mathbf{K}_1\mathbf{J}(t,m_1,m_2-1,k)\\ &\quad + \sum_{k=0}^{n}\mathcal{P}_*(D_0+B_0\leq t, P(\{0\},t)=n-k)\mathbf{K}_1\mathbf{J}(t,m_1,m_2,k).\end{aligned}$$

The recursion simplifies when $m_1=0$ or $m_2=0$ as one or more terms on the right hand side are equal to 0.

Our second main result concerns the corresponding probability generating function. Recall (3) and note that $\hat{\mathbf{J}}(0,z_1,z_2,z_3) = \mathbf{I}$. Define the function $h := [0,\infty)\times\mathbb{D}\times\mathbb{D}\times\mathbb{D}\to\mathbb{C}$ as

$$h(t,z_1,z_2,z_3) := \mathcal{E}_*\left[z_1^{\mathbb{1}_{\{D_0>t\}}} z_2^{\mathbb{1}_{\{D_0\leq t, D_0+B_0>t\}}} z_3^{P(\{0\},t)}\right]. \tag{5}$$

Later, in Lemma 3.1, we will show how this h function can be expressed in terms of the model primitives.

We can associate with each square matrix $\mathbf{M}$ its *matrix exponential* $\mathbf{e}^{\mathbf{M}}$, which is defined as

$$\mathbf{e}^{\mathbf{M}} := \sum_{m=0}^{\infty}\frac{1}{m!}\mathbf{M}^m,$$

with $\mathbf{M}^0 \equiv \mathbf{I}$. It is well-known that $\mathbf{e}^{\mathbf{M}}$ is a well-defined square matrix.

Theorem 2.5. *Assume* (A1)-(A6). *Then for each $z_1,z_2,z_3 \in \mathbb{D}$:*
For Lebesgue almost every $t>0$,

$$\frac{\partial}{\partial t}\hat{\mathbf{J}}(t,z_1,z_2,z_3) = (\mathbf{K}_0 + h(t,z_1,z_2,z_3)\mathbf{K}_1)\hat{\mathbf{J}}(t,z_1,z_2,z_3). \tag{6}$$

Furthermore, $\hat{\mathbf{J}}_0(t,z_1,z_2,z_3) = \mathbf{e}^{\mathbf{K}_0 t}$, and for each integer $n\geq 0$,

$$\hat{\mathbf{J}}_{n+1}(t,z_1,z_2,z_3) = \mathbf{e}^{\mathbf{K}_0 t} + \int_0^t \mathbf{e}^{\mathbf{K}_0 s}h(t-s,z_1,z_2,z_3)\mathbf{K}_1\hat{\mathbf{J}}_n(t-s,z_1,z_2,z_3)ds. \tag{7}$$

Equation (6) from Theorem 2.5 is analogous in form to a result in Ramaswami and Neuts [43] for infinite-server queues where customers arrive to the queue in accordance with a phase-type renewal process. We can calculate $\hat{\mathbf{J}}(t, z_1, z_2, z_3)$ numerically by either numerically solving (6), or by approximating $\hat{\mathbf{J}}(t, z_1, z_2, z_3)$ with $\hat{\mathbf{J}}_n(t, z_1, z_2, z_3)$ for a sufficiently large integer n, and this matrix can be calculated numerically using (7). An extension of the results found in [43] to the case where batches of arrivals occur in accordance with a Markovian arrival process can be found in Masuyama and Takine [39].

We next address the problem of calculating the $\mathbf{C}_{m_1,m_2,n}(t)$ matrices, for each real $t \geq 0$ and each triple of integers $m_1, m_2, n \geq 0$. Theoretically these moments can be derived from their joint PGF by taking derivatives. We present a direct proof via point process theory which yields a recursive scheme that in principle can be used to find all moments. Define $q_{j,k} : [0, \infty) \to [0, \infty)$, for each $j \in \{0, 1\}$ and each integer $k \geq 0$ as follows:

$$q_{j,k}(t) := \mathcal{E}_*\left[\mathbb{1}^j_{\{D_0 \leq t, D_0+B_0>t\}} P(\{0\}, t)^k\right].$$

We follow the convention throughout that empty sums are equal to zero.

Theorem 2.6. *Assume* (A1)-(A6). *Then the matrices* $\mathbf{C}_{m_1,m_2,n}(t)$ *satisfy the recursion*

$$\begin{aligned}\mathbf{C}_{m_1,m_2,n}(t) = \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)t}\mathbb{1}_{\{m_1=m_2=n=0\}} &+ \sum_{j_1=0}^{m_1-1}\binom{m_1}{j_1}\int_0^t \overline{E}_1(s)\mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}\mathbf{K}_1\mathbf{C}_{j_1,m_2,n}(s)ds \\ &+ \sum_{j_2=0}^{m_2-1}\sum_{k=0}^{n}\binom{m_2}{j_2}\binom{n}{k}\int_0^t q_{1,n-k}(s)\mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}\mathbf{K}_1\mathbf{C}_{m,j_2,k}(s)ds \\ &+ \sum_{k=0}^{n-1}\binom{n}{k}\int_0^t q_{0,n-k}(s)\mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}\mathbf{K}_1\mathbf{C}_{m_1,m_2,k}(s)ds.\end{aligned}$$

In particular, the matrices $\mathbf{C}_{1,0,0}(t)$, $\mathbf{C}_{0,1,0}(t)$ *and* $\mathbf{C}_{0,0,1}(t)$ *are as follows*:

$$\begin{aligned}\mathbf{C}_{1,0,0}(t) &= \int_0^t \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}\overline{E}_1(s)\mathbf{K}_1\mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)s}ds \\ \mathbf{C}_{0,1,0}(t) &= \int_0^t \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}(\overline{E_1 \star L_1}(s) - \overline{E_1}(s))\mathbf{K}_1\mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)s}ds \\ \mathbf{C}_{0,0,1}(t) &= \int_0^t \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}q_{0,1}(s)\mathbf{K}_1\mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)s}ds.\end{aligned}$$

These 'moment matrices' satisfy a simpler recursive structure than the matrices we studied in Theorems 2.4 and 2.5, thanks to $(\mathbf{K}_0 + \mathbf{K}_1)t$ and its derivative with respect to t being commutative with respect to matrix multiplication.

Note further that while our main results address the case $M_{\text{nas}}(0) = M(0) = P(0) = 0$, analogous expressions can be derived for arbitrary $M_{\text{nas}}(0)$, $M(0)$, and $P(0)$ if we are given distributions for the remaining elongation periods or lifetimes for each molecule present at time 0. We leave this as an exercise for the interested reader.

Note that other models exhibiting similar dynamics can be analyzed with the same technique. For example, using this technique various time-dependent moments of nascent mRNA and mature mRNA can be derived for the model from [15]. In this model, nascent mRNA must progress through L

independent, identically distributed exponential phases with rate $k+d$, where at the end of phase i, $1 \le i \le L-1$, the nascent mRNA either moves to phase $i+1$ with probability $k/(k+d)$, or dies with probability $d/(k+d)$. At the end of phase L, the nascent mRNA molecule becomes a mature mRNA molecule with probability $k/(k+d)$, or it dies with probability $d/(k+d)$. Finally, the lifetime of each mature mRNA molecule is assumed to be exponentially distributed with rate d_m. Technically, the model from [15] considers only the case where the gene switches back and forth between an active and an inactive state, but we will allow the gene to behave in accordance with an arbitrary MAP.

For each $i \in \{1, 2, \ldots, L\}$, let $M_i(t)$ denote the number of nascent mRNA molecules experiencing phase i at time t, and let $M_{L+1}(t)$ denote the number of mature mRNA molecules present at time t. Next, define the matrices $\mathbf{C}_{m_1,m_2,\ldots,m_{L+1}}(t)$ as

$$\mathbf{C}_{m_1,m_2,\ldots,m_{L+1}}(t) := \left[\mathbb{E}_x\left[\mathbb{1}_{\{I(t)=y\}}\prod_{\ell=1}^{L+1} M_\ell(t)^{m_\ell}\right]\right]_{x,y\in S}, \quad m_1,\ldots,m_{L+1} \in \mathbb{Z}_+.$$

Let $\{Z(t); t \ge 0\}$ denote a finite-state CTMC, having state space $\{1, 2, \ldots, L, L+1, L+2\}$ and transition rate matrix $\mathbf{R}$, where

$$\mathbf{R} = \begin{pmatrix} -(k+d) & k & 0 & 0 & \cdots & 0 & d \\ 0 & -(k+d) & k & 0 & \cdots & 0 & d \\ 0 & 0 & \ddots & \ddots & \cdots & 0 & d \\ \vdots & \vdots & \ddots & 0 & -(k+d) & k & d \\ 0 & 0 & 0 & \cdots & 0 & -d_m & d_m \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}$$

Using the same proof technique, the following result can be established.

Theorem 2.7. *The matrices* $\mathbf{C}_{m_1,m_2,\ldots,m_{L+1}}(t)$ *satisfy the recursion*

$$\mathbf{C}_{m_1,m_2,\ldots,m_{L+1}}(t) = \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)t}\mathbb{1}_{\{m_1=m_2=\cdots=m_{L+1}=0\}} + \sum_{\ell=1}^{L+1}\sum_{j_\ell=0}^{m_\ell-1}\binom{m_\ell}{j_\ell}\int_0^t \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}\mathbb{P}_1(Z(s)=\ell)\mathbf{K}_1\mathbf{C}_{m_1,\ldots,m_{\ell-1},j_\ell,m_{\ell+1},\ldots,m_{L+1}}(s)ds,$$

Note that the probabilities $\mathbb{P}_1(Z(s)=\ell)$ can be expressed in closed-form: for $1 \le \ell \le L$,

$$\mathbb{P}_1(Z(s)=\ell) = \left(\frac{k}{k+d}\right)^{\ell-1}\frac{((k+d)s)^{\ell-1}e^{-(k+d)s}}{(\ell-1)!}$$

and (assuming that $k+d \ne d_m$),

$$\mathbb{P}_1(Z(s)=L+1) = \left(\frac{k}{k+d-d_m}\right)^{L}\left[e^{-d_m s} - \sum_{\ell=0}^{L-1}\frac{((k+d-d_m)s)^\ell e^{-(k+d)s}}{\ell!}\right].$$

We omit the derivation of these probabilities, as they can be calculated using standard methods.

2.3 Numerical Results

Our main results can be used to numerically approximate the time-dependent moments of the numbers of nascent mRNA, mRNA, and proteins. In order to calculate these quantities, we need an efficient method for calculating the $q_{j,k}$ functions, and in general we do not expect to be able to express each $q_{j,k}$ function in closed-form because each such function is in terms of integrals of tail probabilities associated with generally distributed random variables. What we can do though is derive computable expressions when elongation times and lifetimes are discrete.

The moment recursion featured in Theorem 2.6 exhibits a form that makes it highly amenable to numerical work. Fix a real number $T > 0$, and construct a partition of equally-spaced points $\{u_i\}_{i=0}^{p}$ of $[0, T]$, where $u_i = i\delta$, and $\delta = T/p$. Using this partition, we can use Theorem 2.6 to derive the following approximate recursion, which follows from a naive Riemann sum approximation. For $m_1, m_2, n \in \mathbb{Z}_+$ satisfying $\max(m_1, m_2, n) \geq 1$,

$$\begin{aligned}\mathbf{C}_{m_1,m_2,n}(u_1) \approx \delta \sum_{j_1=0}^{m_1-1} \binom{m_1}{j_1} \overline{E}_1(u_1)\mathbf{K}_1\mathbf{C}_{j_1,m_2,n}(u_1) + \delta \sum_{j_2=0}^{m_2-1}\sum_{k=0}^{n} \binom{m_2}{j_2}\binom{n}{k} q_{1,n-k}(u_1)\mathbf{K}_1\mathbf{C}_{m_1,j_2,k}(u_1) \\ + \delta \sum_{k=0}^{n-1} \binom{n}{k} q_{0,n-k}(u_1)\mathbf{K}_1\mathbf{C}_{m_1,m_2,k}(u_1)\end{aligned}$$

and for $1 \leq \ell \leq p-1$,

$$\begin{aligned}\mathbf{C}_{m_1,m_2,n}(u_{\ell+1}) \approx \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)\delta}\mathbf{C}_{m_1,m_2,n}(u_\ell) + \delta \sum_{j_1=0}^{m_1-1} \binom{m_1}{j_1} \overline{E}_1(u_{\ell+1})\mathbf{K}_1\mathbf{C}_{j_1,m_2,n}(u_{\ell+1}) \\ + \delta \sum_{j_2=0}^{m_2-1}\sum_{k=0}^{n} \binom{m_2}{j_2}\binom{n}{k} q_{1,n-k}(u_{\ell+1})\mathbf{K}_1\mathbf{C}_{m_1,j_2,k}(u_{\ell+1}) + \delta \sum_{k=0}^{n-1} \binom{n}{k} q_{0,n-k}(u_{\ell+1})\mathbf{K}_1\mathbf{C}_{m_1,m_2,k}(u_{\ell+1})\end{aligned}$$

where we again recall that $\mathbf{C}_{0,0,0}(u_\ell) = \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)u_\ell}$. This simple recursion seems to work quite well when m_1, m_2, and n are relatively small. We will use this recursion within an example at the end of Section 4.

3 Proofs of the main results

In this Section we prove Theorems 2.4-2.6. We will use two tools from the theory of Poisson processes (more generally point processes) which will be stated below. In the meantime we present a lemma that quantifies the relationship between D_0, B_0, and $P(\{0\}, t)$. In order to simplify the statement of this lemma, we define the function $r : \mathbb{R}_+^2 \to \mathbb{R}$ as

$$r(a, b) := k_2 \int_a^b (\overline{E_2 \star L_2}(t-x) - \overline{E_2}(t-x))dx.$$

Henceforth, by convention "Poisson with mean 0" means almost surely equal to 0.

Lemma 3.1. *The function* $h(t, z_1, z_2, z_3) := \mathcal{E}_*\left[z_1^{\mathbb{1}_{\{D_0>t\}}} z_2^{\mathbb{1}_{\{D_0\le t, D_0+B_0>t\}}} z_3^{P(\{0\},t)}\right]$ *satisfies*

$$h(t, z_1, z_2, z_3) = z_1\overline{E_1}(t) + z_2\int_{[0,t]} \overline{L_1}(t-u)e^{-(1-z_3)r(u,t)}dE_1(u) + \int_{[0,t]}\int_{[0,t-u]} e^{-(1-z_3)r(u,u+v)}dL_1(v)dE_1(u). \tag{8}$$

Moreover, the following statements are true:

$$\mathcal{P}_*\big(D_0+B_0\le t, P(\{0\},t)=k\big) = \int_{[0,t]}\int_{[0,t-u]} \frac{\big(r(u,u+v)\big)^k e^{-r(u,u+v)}}{k!}dL_1(v)dE_1(u), \tag{9}$$

$$\mathcal{P}_*\big(D_0\le t, D_0+B_0>t, P(\{0\},t)=k\big) = \int_{[0,t]} \overline{L_1}(t-u)\frac{\big(r(u,t)\big)^k e^{-r(u,t)}}{k!}dE_1(u), \tag{10}$$

and the function $q_{j,k}(t) := \mathcal{E}_*\left[\mathbb{1}^j_{\{D_0\le t, D_0+B_0>t\}} P(\{0\},t)^k\right]$ *satisfies*

$$q_{j,k}(t) = \int_{[0,\infty)}\int_{[0,\infty)} \big(\mathbb{1}_{\{u\le t, u+v>t\}}\big)^j \mu_k\big((u\wedge t, (u+v)\wedge t]\big)dL_1(v)dE_1(u), \tag{11}$$

where $\mu_k((a,b])$ *is the* k*th moment of a Poisson random variable having mean* $r(a,b)$.

Proof. First, note that a protein initiated at time x in the $\mathcal{P}_*$-setting is present at time t if its elongation time ($\sim E_2$) is at most $t-x$, and its elongation time plus lifetime ($\sim E_2 \star L_2$) is greater than $t-x$. Thus, conditional on both D_0 and B_0, $P(\{0\},t)$ is a Poisson random variable with mean $\lambda_t(D_0,B_0)$, where

$$\lambda_t(a,b) := k_2\int_{\min(a,t)}^{\min(a+b,t)} (\overline{E_2\star L_2}(t-x) - \overline{E_2}(t-x))dx = r\big(\min(a,t), \min(a+b,t)\big).$$

Thus, for functions f_1, f_2, where $f_1 : [0,\infty)\times[0,\infty)\to[0,\infty)$ is $\mathcal{B}([0,\infty))\otimes\mathcal{B}([0,\infty))$-measurable, and $f_2 : \mathbb{Z}_+\to[0,\infty)$, by conditioning on both B_0 and D_0, we can write

$$\begin{aligned}\mathcal{E}_*\big[f_1(D_0,B_0)f_2(P(\{0\},t))\big] &= \mathcal{E}_*\Big[f_1(D_0,B_0)\mathcal{E}_*\big[f_2(P(\{0\},t))\big|D_0,B_0\big]\Big]\\ &= \int_0^\infty\int_0^\infty f_1(u,v)\phi_2(t,u,v)dL_1(v)dE_1(u),\end{aligned}$$

where $\phi_2(t,u,v)$ is the expected value of $f_2(N^*)$, where $N^*\sim\text{Pois}(\lambda_t(u,v))$.

From this observation Formulas (8), (9), (10), and (11) can be derived quickly, so we will only spend time establishing (8): since $\{D_0>t\}$ and $\{D_0+B_0\le t\}$ are disjoint events we have

$$\mathcal{E}_*\big[z_1^{\mathbb{1}_{\{D_0>t\}}} z_2^{\mathbb{1}_{\{D_0\le t, D_0+B_0>t\}}} z_3^{P(\{0\},t)}\big] = \mathcal{E}_*\big[(z_1\mathbb{1}_{\{D_0>t\}} + z_2\mathbb{1}_{\{D_0\le t, D_0+B_0>t\}} + \mathbb{1}_{\{D_0+B_0\le t\}})z_3^{P(\{0\},t)}\big].$$

Condition on D_0, B_0 as above, set $f_2(n) = z_3^n$ and e.g. $f_1(u,v) = z_1\mathbb{1}_{\{u>t\}}$ for the first term in the $\mathcal{P}_*$ expectation, and recall that the probability generating function for a Poisson(λ) random variable is $\phi(z) = e^{-\lambda(1-z)}$ to see that this is equal to

$$\begin{aligned}&\int_0^\infty\int_0^\infty [z_1\mathbb{1}_{\{u>t\}} + z_2\mathbb{1}_{\{u\le t, u+v>t\}} + \mathbb{1}_{\{u+v\le t\}}]e^{-(1-z_3)r(u\wedge t, u+v\wedge t)}dL_1(v)dE_1(u)\\ &= z_1\overline{E_1}(t) + z_2\int_{[0,t]}\overline{L_1}(t-u)e^{-(1-z_3)r(u,t)}dE_1(u) \quad + \int_{[0,t]}\int_{[0,t-u]} e^{-(1-z_3)r(u,u+v)}dL_1(v)dE_1(u),\end{aligned}$$

which proves the claim (8). □

We now introduce the two tools from the theory of point processes that we will apply in the proofs. Note that the treatment below serves only to give a rudimentary understanding of these results, sufficient to be comfortable with our application of them. Readers interested in technical details, proofs and/or generalised statements are referred to one of many text books on the subject, e.g. [4, 34].

A Poisson process $\mathcal{A} := \{\mathcal{A}(t); t \geq 0\}$ with rate $k > 0$ is a counting process that can also be viewed as a random measure on the Borel sets $\mathcal{B}$ of $[0, \infty)$, with $\mathcal{A}(B)$ denoting the number of points in $B \subset [0, \infty)$. Note that for $s \leq t$, $\mathbb{E}[\mathcal{A}([s, t])] = k(t - s)$. The general theory of point processes provides a family of non-random probability measures $\{\mathcal{P}_s; s \geq 0\}$ on $(\Omega, \mathcal{F})$ (with corresponding expectations $\{\mathcal{E}_s; s \geq 0\}$), called *Palm probabilities*, for which $\mathcal{P}_s(A)$ can be interpreted as the probability of the event A given that the process $\mathcal{A}$ has a point at s. These measures are non-trivial since the event that $\mathcal{A}$ has a point at s has probability 0. Note that we have already informally encountered such a probability measure, namely $\mathcal{P}_0$, which we introduced as a probability measure under which there was an mRNA binding at time 0, and we calculated expectations of quantities (that didn't depend any further on the Poisson process) arising from that binding.

For $s \geq 0$, let δ_s denote the measure on $\mathcal{B}$ that puts a unit mass at the point s (i.e. $\delta_s(B) = \mathbb{1}_{\{s \in B\}}$). Given a Poisson process $\mathcal{A}$, let $\mathcal{A}_s := \{\mathcal{A}_s(t); t \geq 0\}$ denote the point process that is equal to $\mathcal{A}$ except that $\mathcal{A}_s$ contains a point at s. The following result, attributed to Silvnyak and Mecke, can be found on page 130 of [4]. It says that conditioning a Poisson process on having a point at time s (we reiterate that this is an event of probability 0) is equivalent to adding a point at time s to the unconditioned process.

Theorem 3.1. *Let $\mathcal{A}$ be a Poisson process. Then for Lebesgue a.e. s, the law of $\mathcal{A}$ under the Palm measure $\mathcal{P}_s$ is equal to the law of $\mathcal{A}_s$ under $\mathbb{P}$.*

For example, for an interval I, according to Theorem 3.1,

$$\mathcal{P}_s(\mathcal{A}(I) = n) = \mathbb{P}(\mathcal{A}_s(I) = n) = \begin{cases} \mathbb{P}(\mathcal{A}(I) = n), & \text{if } s \notin I, \\ \mathbb{P}(\mathcal{A}(I) = n - 1), & \text{if } s \in I. \end{cases}$$

Let $X := \{X(s); s \geq 0\}$ be a non-negative random process defined on $(\Omega, \mathcal{F})$. The following is a special case of the *Campbell-Mecke Formula* [4] for Poisson processes. It gives an expression for the expectation (under $\mathbb{P}$) of the sum of the values of the $X(\cdot)$ process at the points of the Poisson process up to time t, in terms of the time average (up to t) of the Palm expectations.

Theorem 3.2. (*Campbell-Mecke Formula*) *Let $\mathcal{A}$ be a Poisson process with rate k and X be a non-negative (and $\mathcal{F} \otimes \mathcal{B}$-measurable) process. Then for each $t \geq 0$,*

$$\mathbb{E}\left[\int_0^t X(s)\mathcal{A}(ds)\right] = \int_0^t \mathcal{E}_s[X(s)]k\,ds. \tag{12}$$

In the special case where X is non-random (with $X(s) = x(s)$ for each s), both sides of (12) are equal to $\int_0^t x(s)k\,ds$. In the special case where $X(s) = \mathbb{1}_{\{\mathcal{A}(I)=n\}}$ for some interval I, we have that the left hand side of (12) is equal to

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{\{\mathcal{A}(I)=n\}}\mathcal{A}([0,t])] &= \mathbb{E}[\mathcal{A}([0,t])|\mathcal{A}(I) = n]\mathbb{P}(\mathcal{A}(I) = n) \\ &= \Big(\mathbb{E}[\mathcal{A}([0,t] \setminus I)|\mathcal{A}(I) = n] + \mathbb{E}[\mathcal{A}([0,t] \cap I)|\mathcal{A}(I) = n]\Big)\mathbb{P}(\mathcal{A}(I) = n) \\ &= \left(k|[0,t] \setminus I| + n\frac{|[0,t] \cap I|}{|I|}\right)\mathbb{P}(\mathcal{A}(I) = n), \end{aligned} \tag{13}$$

where $|B|$ denotes the Lebesgue measure of B. The right hand side of (12) is equal to

$$\begin{aligned}\int_0^t \mathcal{E}_s(\mathcal{A}(I)=n)k\,ds &= k\int_{[0,t]\setminus I}\mathbb{P}(\mathcal{A}(I)=n)ds + k\int_{[0,t]\cap I}\mathbb{P}(\mathcal{A}(I)=n-1)ds \\ &= k\mathbb{P}(\mathcal{A}(I)=n)|[0,t]\setminus I| + k\mathbb{P}(\mathcal{A}(I)=n-1)|[0,t]\cap I|, \end{aligned} \tag{14}$$

where we have used Theorem 3.1 in the first line. It is now an elementary exercise to verify that (13) and (14) are equal.

We now have the tools we need to prove our main results.

We define the conception time of any mRNA or protein molecule as the time of the binding of the RNA polymerase (that led to the creation of the molecule) with the (active) gene. Throughout our derivations, for each $t \geq 0$, and each $B \in \mathcal{B}([0,\infty))$, $M(B,t)$ (resp. $P(B,t)$, $M_{\text{nas}}(B,t)$) represents the number of mRNA (resp. protein, nascent mRNA) molecules present at time t with conception time in B. The elongation time and lifetime associated to the first arrival in the MAP are denoted by $D^{(1)}$ and $B^{(1)}$ respectively. In each case our proof starts with the observation that (for particular choices of f_1, f_2, and f_3) we have

$$\begin{aligned}
&f_1(M_{\text{nas}}(t))f_2(M(t))f_3(P(t)) \\
&= f_1(0)f_2(0)f_3(0)\mathbb{1}_{\{N(t)=0\}} \\
&\quad + \sum_{z,w\in S}\int_{(0,t]} \mathbb{1}_{\{N(s-)=0,I(s-)=z\}}\mathbb{1}_{\{D^{(1)}>t-s\}} \\
&\qquad\qquad \times f_1\big(1+M_{\text{nas}}((s,t],t)\big)f_2\big(M((s,t],t)\big)f_3\big(P((s,t],t)\big)A^{(1)}_{z,w}(ds) \\
&\quad + \sum_{z,w\in S}\int_{(0,t]} \mathbb{1}_{\{N(s-)=0,I(s-)=z\}}\mathbb{1}_{\{D^{(1)}\leq t-s, D^{(1)}+B^{(1)}>t-s\}} \\
&\qquad\qquad f_1\big(M_{\text{nas}}((s,t],t)\big)f_2\big(1+M((s,t],t)\big)f_3\big(P(\{s\},t)+P((s,t],t)\big)A^{(1)}_{z,w}(ds) \\
&\quad + \sum_{z,w\in S}\int_{(0,t]} \mathbb{1}_{\{N(s-)=0,I(s-)=z\}}\mathbb{1}_{\{D^{(1)}+B^{(1)}\leq t-s\}} \\
&\qquad\qquad f_1\big(M_{\text{nas}}((s,t],t)\big)f_2\big(M((s,t],t)\big)f_3\big(P(\{s\},t)+P((s,t],t)\big)A^{(1)}_{z,w}(ds).
\end{aligned} \tag{15}$$

The first term on the right hand side of (15) is the case that there is no conception by time t, while in the remaining terms the first conception time is $s \in (0,t]$. Depending on $D^{(1)}, B^{(1)}$ this mRNA conceived at time s might be nascent at time t (which contributes to $M_{\text{nas}}(t)$, and this corresponds to the second term in (15)), or has finished its elongation period by time t, and in this case it may be still alive (see the third term in (15)) or not (see the final term in (15)) at time t.

We then multiply both sides by $\mathbb{1}_{\{I(t)=y\}}$ and use the particular form of f_1, f_2, f_3 to write $f_i(n+m)$ in terms of $f_i(k)$ etc. (in our first two proofs, each f_i factorizes, and in our third proof we use the binomial expansion). We take expectations of both sides and apply both Theorems 3.2 and 3.1 to the right hand side. Exploiting independence allows us to factorize the resulting Palm expectation. Simplifying, expressing in matrix form, and differentiating with respect to t completes the proof.

Proof of Theorem 2.4. Let $t > 0$, $m_1, m_2, n \in \mathbb{Z}_+$. Using (15) with $f_1(m) = \mathbb{1}_{\{m=m_1\}}$, $f_2(m) = \mathbb{1}_{\{m=m_2\}}$ and $f_3(m) = \mathbb{1}_{\{m=n\}}$ we obtain,

$$
\begin{aligned}
&\mathbb{1}_{\{M_{\text{nas}}(t)=m_1, M(t)=m_2, P(t)=n, I(t)=y\}} \\
&= \mathbb{1}_{\{N(t)=0, I(t)=y\}} \mathbb{1}_{\{m_1=m_2=n=0\}} \\
&\quad + \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}>t-s\}} \\
&\qquad\qquad \times \mathbb{1}_{\{M_{\text{nas}}((s,t],t)=m_1-1, M((s,t],t)=m_2, P((s,t],t)=n, I(t)=y\}} A^{(1)}_{z,w}(ds) \\
&\quad + \sum_{k=0}^{n} \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}\le t-s, D^{(1)}+B^{(1)}>t-s\}} \mathbb{1}_{\{P(\{s\},t)=n-k\}} \\
&\qquad\qquad \times \mathbb{1}_{\{M_{\text{nas}}((s,t],t)=m_1, M((s,t],t)=m_2-1, P((s,t],t)=k, I(t)=y\}} A^{(1)}_{z,w}(ds) \\
&\quad + \sum_{k=0}^{n} \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}+B^{(1)}\le t-s\}} \mathbb{1}_{\{P(\{s\},t)=n-k\}} \\
&\qquad\qquad \times \mathbb{1}_{\{M_{\text{nas}}((s,t],t)=m_1, M((s,t],t)=m_2, P((s,t],t)=k, I(t)=y\}} A^{(1)}_{z,w}(ds).
\end{aligned} \tag{16}
$$

Note that if e.g. $m_1 = 0$ then $\mathbb{1}_{\{M_{\text{nas}}((s,t],t)=m_1-1\}} = 0$ above. After taking expectations of both sides, while applying the Campbell-Mecke formula to e.g. the process

$$
X_{t,k}(s) := \mathbb{1}_{\{N(s-)=0, I(s-)=z, D^{(1)}+B^{(1)}\le t-s, P(\{s\},t)=n-k\}} \mathbb{1}_{\{M_{\text{nas}}((s,t],t)=0, M((s,t],t)=0, P((s,t],t)=k, I(t)=y\}},
$$

for the last term on the right-hand-side (and similar processes for the other terms on the right hand side) of (16), we find that for each $x \in S$, and $t \ge 0$,

$$
\begin{aligned}
&\mathbb{P}_x(M_{\text{nas}}(t) = m_1, M(t) = m_2, P(t) = n, I(t) = y) \\
&= \mathbb{P}_x(N(t) = 0, I(t) = y) \mathbb{1}_{\{m_1=m_2=n=0\}} \\
&\quad + \sum_{z,w\in S} \int_0^t \mathcal{P}^{(z,w)}_{x,s} \Big(N(s-) = 0, I(s-) = z, D^{(1)} > t-s, I(t) = y, \\
&\qquad\qquad M_{\text{nas}}((s,t],t) = m_1 - 1, M((s,t],t) = m_2, P((s,t],t) = n \Big) k_1(z,w) ds \\
&\quad + \sum_{k=0}^{n} \sum_{z,w\in S} \int_0^t \mathcal{P}^{(z,w)}_{x,s} \Big(N(s-) = 0, I(s-) = z, D^{(1)} \le t-s, D^{(1)} + B^{(1)} > t-s, P(\{s\},t) = n-k, \\
&\qquad\qquad M_{\text{nas}}((s,t],t) = m_1, M((s,t],t) = m_2 - 1, P((s,t],t) = k, I(t) = y \Big) ds \\
&\quad + \sum_{k=0}^{n} \sum_{z,w\in S} \int_{(0,t]} \mathcal{P}^{(z,w)}_{x,s} \Big(N(s-) = 0, I(s-) = z, D^{(1)} + B^{(1)} \le t-s, \\
&\qquad\qquad P(\{s\},t) = n-k, M_{\text{nas}}((s,t],t) = m_1, M((s,t],t) = m_2,
\end{aligned} \tag{17}
$$

$$P((s,t],t) = k, I(t) = y\Big)k_1(z,w)ds,$$

where $\{\mathcal{P}^{(z,w)}_{x,s}\}_{s\geq 0}$ denotes the family of Palm distributions associated with the Poisson process $A^{(1)}_{z,w}$ (when the underlying probability measure is $\mathbb{P}_x$). Next, observe e.g. in the last term that the event $\{N(s-) = 0, I(s-) = z\}$ depends only on the behaviour of all points on the interval $(0,s)$, the event $\{M_{\text{nas}}((s,t],t) = m_1, M((s,t],t) = m_2, P((s,t],t) = k, I(t) = y\}$ depends only on the behaviour of the point processes in the set $(s,t]$ (while simultaneously keeping in mind that $I(s) = w$) and the event $\{D^{(1)} + B^{(1)} \leq t - s\}$ does not depend on the point process at all for fixed s.

Applying now Theorem 3.1 yields, for each $t \geq 0$ and Lebesgue almost-all $s \in [0,t]$, the probability in the last term on the right hand side of (17) is

$$\begin{aligned}
&\mathcal{P}^{(z,w)}_{x,s}\Big(N(s-) = 0, I(s-) = z, D^{(1)} + B^{(1)} \leq t - s,\\
&\qquad P(\{s\},t) = n-k, M_{\text{nas}}((s,t],t) = m_1, M((s,t],t) = m_2, P((s,t],t) = k, I(t) = y\Big)\\
&= \mathbb{P}_x\big(N(s) = 0, I(s) = z\big)\mathcal{P}_*\big(D_0 + B_0 \leq t - s, P(\{0\}, t-s) = n-k\big)\\
&\quad \times \mathbb{P}_w\big(M_{\text{nas}}(t-s) = m_1, M(t-s) = m_2, P(t-s) = k, I(t-s) = y\big).
\end{aligned}$$

The other terms are handled similarly. Hence,

$$\begin{aligned}
&\mathbb{P}_x(M_{\text{nas}}(t) = m_1, M(t) = m_2, P(t) = n, I(t) = y)\\
&= \mathbb{P}_x(N(t) = 0, I(t) = y)\mathbb{1}_{\{m_1=m_2=n=0\}}\\
&\quad + \sum_{z,w\in S}\int_0^t \mathbb{P}_x\big(N(s) = 0, I(s) = z\big)\mathcal{P}_*\big(D_0 > t-s, P(\{0\}, t-s) = n-k\big)k_1(z,w)\\
&\qquad\qquad \times \mathbb{P}_w\big(M_{\text{nas}}(t-s) = m_1 - 1, M(t-s) = m_2, P(t-s) = n, I(t-s) = y\big)ds\\
&\quad + \sum_{k=0}^{n}\sum_{z,w\in S}\int_0^t \mathbb{P}_x\big(N(s) = 0, I(s) = z\big)\mathcal{P}_*\big(D_0 \leq t-s, D_0 + B_0 > t-s, P(\{0\}, t-s) = n-k\big)k_1(z,w)\\
&\qquad\qquad \times \mathbb{P}_w\big(M_{\text{nas}}(t-s) = m_1, M(t-s) = m_2 - 1, P(t-s) = k, I(t-s) = y\big)ds\\
&\quad + \sum_{k=0}^{n}\sum_{z,w\in S}\int_0^t \mathbb{P}_x\big(N(s) = 0, I(s) = z\big)\mathcal{P}_*\big(D_0 + B_0 \leq t-s, P(\{0\}, t-s) = n-k\big)\\
&\qquad\qquad \mathbb{P}_w\big(M_{\text{nas}}(t-s) = m_1, M(t-s) = m_2, P(t-s) = k, I(t-s) = y\big)ds.
\end{aligned}$$

Observe also that from the Kolmogorov forward equations associated with the CTMC $\{(N(t), I(t)); t \geq 0\}$, the probability $\mathbb{P}_x(N(t) = 0, I(t) = y)$ is the element found in row x, column y of the matrix exponential $\mathbf{e}^{\mathbf{K}_0 t}$. Expressing the above system in matrix form we obtain

$$\begin{aligned}
\mathbf{J}(t, m_1, m_2, n) &= \mathbf{e}^{\mathbf{K}_0 t}\mathbb{1}_{\{m_1=m_2=n=0\}}\\
&\quad + \int_0^t \mathcal{P}_*\big(D_0 > t-s, P(\{0\}, t-s) = n-k\big)\mathbf{e}^{\mathbf{K}_0 s}\mathbf{K}_1\mathbf{J}(t-s, m_1-1, m_2, k)ds
\end{aligned}$$

$$+\sum_{k=0}^{n}\int_0^t \mathcal{P}_*\big(D_0 \le t-s, D_0+B_0 > t-s, P(\{0\},t-s)=n-k\big)$$
$$\times \mathbf{e}^{\mathbf{K}_0 s}\mathbf{K}_1\mathbf{J}(t-s,m_1,m_2-1,k)ds$$
$$+\sum_{k=0}^{n}\int_0^t \mathcal{P}_*\big(D_0+B_0 \le t-s, P(\{0\},t-s)=n-k\big)\mathbf{e}^{\mathbf{K}_0 s}\mathbf{K}_1\mathbf{J}(t-s,m_1,m_2,k)ds.$$

Using the change of variables $u = t - s$ this becomes

$$\mathbf{J}(t,m_1,m_2,n) = \mathbf{e}^{\mathbf{K}_0 t}\Big[\mathbb{1}_{\{m_1=m_2=n=0\}} + \int_0^t \mathcal{P}_*\big(D_0 > u, P(\{0\},u)=n-k\big)\mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{J}(u,m_1-1,m_2,k)du$$
$$+\sum_{k=0}^{n}\int_0^t \mathcal{P}_*\big(D_0 \le u, D_0+B_0 > u, P(\{0\},u)=n-k\big)$$
$$\times \mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{J}(u,m_1,m_2-1,k)du$$
$$+\sum_{k=0}^{n}\int_0^t \mathcal{P}_*\big(D_0+B_0 \le u, P(\{0\},u)=n-k\big)\mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{J}(u,m_1,m_2,k)ds\Big].$$

Multiplying both sides of the equality (on the left) by $\mathbf{e}^{-\mathbf{K}_0 t}$, then taking derivatives of both sides with respect to t establishes the result. □

We now turn our attention to the probability generating function $\hat{\mathbf{J}}$. The idea behind the proof we give for Theorem 2.5 is analogous to the idea we used to derive Theorem 2.4. Again, this appears to be closely related to the argument from pages 68 and 69 of [38] for the PH/G/∞ queue, except there the authors do not think in terms of point processes.

Proof of Theorem 2.5. Observe that for each $y \in S$ and $t > 0$, using (15) with $f_1(m) = z_1^m$, $f_2(m) = z_2^m$, and $f_3(m) = z_3^m$, we obtain

$$z_1^{M_{\text{nas}}(t)} z_2^{M(t)} z_3^{P(t)} \mathbb{1}_{\{I(t)=y\}}$$
$$= \mathbb{1}_{\{N(t)=0,I(t)=y\}} + \sum_{z,w\in S}\int_{(0,t]} \mathbb{1}_{\{N(s-)=0,I(s-)=z\}}\mathbb{1}_{\{D^{(1)}>t-s\}} z_1$$
$$z_1^{M_{\text{nas}}((s,t],t)} z_2^{M((s,t],t)} z_3^{P((s,t],t)} A^{(1)}_{z,w}(ds)$$
$$+\sum_{z,w\in S}\int_{(0,t]} \mathbb{1}_{\{N(s-)=0,I(s-)=z\}}\mathbb{1}_{\{D^{(1)}\le t-s, D^{(1)}+B^{(1)}>t-s\}} z_2 z_3^{P(\{s\},t)}$$
$$z_1^{M_{\text{nas}}((s,t],t)} z_2^{M((s,t],t)} z_3^{P((s,t],t)} A^{(1)}_{z,w}(ds)$$
$$+\sum_{z,w\in S}\int_{(0,t]} \mathbb{1}_{\{N(s-)=0,I(s-)=z\}}\mathbb{1}_{\{D^{(1)}+B^{(1)}\le t-s\}} z_3^{P(\{s\},t)}$$
$$z_1^{M_{\text{nas}}((s,t],t)} z_2^{M((s,t],t)} z_3^{P((s,t],t)} A^{(1)}_{z,w}(ds).$$

Taking expectations, and applying Theorems 3.2 and 3.1 yields, for each $x, y \in S$,

$$\begin{aligned}
&\mathbb{E}_x[z_1^{M_{\text{nas}}(t)} z_2^{M(t)} z_3^{P(t)} \mathbb{1}_{\{I(t)=y\}}] \\
&= \mathbb{P}_x(N(t) = 0, I(t) = y) \\
&\quad + \sum_{z,w \in S} \int_0^t \mathbb{P}_x(N(s) = 0, I(s) = z) k_1(z, w) h(t-s, z_1, z_2, z_3) \mathbb{E}_w[z_1^{M_{\text{nas}}(t-s)} z_2^{M(t-s)} z_3^{P(t-s)} \mathbb{1}_{\{I(t)=y\}}] ds.
\end{aligned}$$

In matrix form this is

$$\begin{aligned}
\hat{\mathbf{J}}(t, z_1, z_2, z_3) &= \mathbf{e}^{\mathbf{K}_0 t} + \int_0^t h(t-s, z_1, z_2, z_3) \mathbf{e}^{\mathbf{K}_0 s} \mathbf{K}_1 \hat{\mathbf{J}}(t-s, z_1, z_2, z_3) ds \\
&= \mathbf{e}^{\mathbf{K}_0 t} \left[\mathbf{I} + \int_0^t h(u, z_1, z_2, z_3) \mathbf{e}^{-\mathbf{K}_0 u} \mathbf{K}_1 \hat{\mathbf{J}}(u, z_1, z_2, z_3) du \right],
\end{aligned}$$

and taking the partial derivative of both sides with respect to t gives

$$\frac{\partial}{\partial t} \hat{\mathbf{J}}(t, z_1, z_2, z_3) = (\mathbf{K}_0 + h(t, z_1, z_2, z_3) \mathbf{K}_1) \hat{\mathbf{J}}(t, z_1, z_2, z_3)$$

proving (6). Finally, note that (7) can be established using a similar idea: indeed, a simple modification of this proof shows that for each integer $n \geq 0$ (where $N((s,t]) := N(t) - N(s)$),

$$\begin{aligned}
&z_1^{M_{\text{nas}}(t)} z_2^{M(t)} z_3^{P(t)} \mathbb{1}_{\{I(t)=y, N(t) \leq n+1\}} \\
&= \mathbb{1}_{\{N(t)=0, I(t)=y\}} + \sum_{z,w \in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)} > t-s\}} z_1 \\
&\qquad\qquad z_1^{M_{\text{nas}}((s,t],t)} z_2^{M((s,t],t)} z_3^{P((s,t],t)} \mathbb{1}_{\{I(t)=y, N((s,t]) \leq n\}} A^{(1)}_{z,w}(ds) \\
&\qquad + \sum_{z,w \in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)} \leq t-s, D^{(1)}+B^{(1)} > t-s\}} z_2 z_3^{P(\{s\},t)} \\
&\qquad\qquad z_1^{M_{\text{nas}}((s,t],t)} z_2^{M((s,t],t)} z_3^{P((s,t],t)} \mathbb{1}_{\{I(t)=y, N((s,t]) \leq n\}} A^{(1)}_{z,w}(ds) \\
&\qquad + \sum_{z,w \in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}+B^{(1)} \leq t-s\}} z_3^{P(\{s\},t)} \\
&\qquad\qquad z_1^{M_{\text{nas}}((s,t],t)} z_2^{M((s,t],t)} z_3^{P((s,t],t)} \mathbb{1}_{\{I(t)=y, N((s,t]) \leq n\}} A^{(1)}_{z,w}(ds).
\end{aligned}$$

Again, after taking expectations, applying Theorems 3.2 and 3.1, and rewriting the resulting equations as a matrix equation, we get

$$\hat{\mathbf{J}}_{n+1}(t, z_1, z_2) = \mathbf{e}^{\mathbf{K}_0 t} + \int_0^t \mathbf{e}^{\mathbf{K}_0 s} h(t-s, z_1, z_2) \mathbf{K}_1 \hat{\mathbf{J}}_n(t-s, z_1, z_2) ds$$

which establishes (7). □

It remains to prove Theorem 2.6.

Proof of Theorem 2.6. Using (15) with $f_1(k) = k^{m_1}$, $f_2(k) = k^{m_2}$, and $f_3(k) = k^n$ gives

$$
\begin{aligned}
&M_{\mathrm{nas}}(t)^{m_1} M(t)^{m_2} P(t)^n \mathbb{1}_{\{I(t)=y\}} \\
&= \mathbb{1}_{\{m_1=m_2=n=0\}} \mathbb{1}_{\{N(t)=0, I(t)=y\}} \\
&\quad + \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}>t-s\}} (1 + M_{\mathrm{nas}}((s,t],t))^{m_1} \\
&\qquad\qquad M((s,t],t)^{m_2} P((s,t],t)^n A^{(1)}_{z,w}(ds) \\
&\quad + \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}\le t-s, D^{(1)}+B^{(1)}>t-s\}} M_{\mathrm{nas}}((s,t],t)^{m_1} \\
&\qquad\qquad (1 + M((s,t],t))^{m_2} (P(\{s\},t) + P((s,t],t))^n A^{(1)}_{z,w}(ds) \\
&\quad + \sum_{z,w\in S} \int_{(0,t]} \mathbb{1}_{\{N(s-)=0, I(s-)=z\}} \mathbb{1}_{\{D^{(1)}+B^{(1)}\le t-s\}} M_{\mathrm{nas}}((s,t],t)^{m_1} M((s,t],t)^{m_2} \\
&\qquad\qquad (P(\{s\},t) + P((s,t],t))^n A^{(1)}_{z,w}(ds).
\end{aligned}
$$

After expanding the integrands with applications of the binomial theorem, taking expectations and applying Theorems 3.2 and 3.1, we observe that for each $x, y \in S$,

$$
\begin{aligned}
&\mathbb{E}_x[M_{\mathrm{nas}}(t)^{m_1} M(t)^{m_2} P(t)^n \mathbb{1}_{\{I(t)=y\}}] \\
&= \mathbb{P}_x(N(t)=0, I(t)=y) \mathbb{1}_{\{m_1=m_2=n=0\}} \\
&\quad + \sum_{z,w\in S} \int_0^t \mathbb{P}_x\big(N(s)=0, I(s)=z\big) \sum_{j_1=0}^{m_1} \binom{m_1}{j_1} \overline{E}_1(t-s) \\
&\qquad\qquad \times k_1(z,w) \mathbb{E}_w\big[M_{\mathrm{nas}}(t-s)^{j_1} M(t-s)^{m_2} P(t-s)^n \mathbb{1}_{\{I(t-s)=y\}}\big] ds \\
&\quad + \sum_{z,w\in S} \int_0^t \mathbb{P}_x\big(N(s)=0, I(s)=z\big) \sum_{j_2=0}^{m_2} \sum_{k=0}^{n} \binom{m_2}{j_2} \binom{n}{k} \mathcal{E}_*\Big[\mathbb{1}_{\{D_0\le t-s, D_0+B_0>t-s\}} P(\{0\}, t-s)^{n-k}\Big] \\
&\qquad\qquad \times k_1(z,w) \mathbb{E}_w\big[M_{\mathrm{nas}}(t-s)^{m_1} M(t-s)^{j_2} P(t-s)^k \mathbb{1}_{\{I(t-s)=y\}}\big] ds \\
&\quad + \sum_{z,w\in S} \int_0^t \mathbb{P}_x\big(N(s)=0, I(s)=z\big) \sum_{k=0}^{n} \binom{n}{k} \mathcal{E}_*\Big[\mathbb{1}_{\{D_0+B_0\le t-s\}} P(\{0\}, t-s)^{n-k}\Big] \\
&\qquad\qquad \times k_1(z,w) \mathbb{E}_w\big[M_{\mathrm{nas}}(t-s)^{m_1} M(t-s)^{m_2} P(t-s)^k \mathbb{1}_{\{I(t-s)=y\}}\big] ds.
\end{aligned}
$$

In matrix form (and following a substitution $u = t - s$) this system can be expressed, after some

algebra, as

$$\begin{aligned}\mathbf{C}_{m_1,m_2,n}(t) = \mathbf{e}^{\mathbf{K}_0 t}\Bigg[&\mathbb{1}_{\{m_1=m_2=n=0\}} + \sum_{j_1=0}^{m_1-1}\binom{m_1}{j_1}\int_0^t \overline{E}_1(s)\mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{C}_{j_1,m_2,n}(u)du \\ &+ \sum_{j_2=0}^{m_2-1}\sum_{k=0}^{n}\binom{m_2}{j_2}\binom{n}{k}\int_0^t q_{1,n-k}(u)\mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{C}_{m_1,j_2,k}(u)du \\ &+ \sum_{k=0}^{n-1}\binom{n}{k}\int_0^t q_{0,n-k}(u)\mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{C}_{m_1,m_2,k}(u)du \\ &+ \int_0^t \mathbf{e}^{-\mathbf{K}_0 u}\mathbf{K}_1\mathbf{C}_{m_1,m_2,n}(u)du\Bigg]\end{aligned}$$

where in our simplifications, we repeatedly make use of the following observation: for each integer $n \geq 1$, and each integer $k \in \{0, 1, \ldots, n-1\}$,

$$\begin{aligned}&\mathcal{E}_*\Big[\mathbb{1}_{\{D_0\leq t-s, D_0+B_0>t-s\}}P(\{0\}, t-s)^{n-k}\Big] + \mathcal{E}_*\Big[\mathbb{1}_{\{D_0+B_0\leq t-s\}}P(\{0\}, t-s)^{n-k}\Big] \\ &= \mathcal{E}_*\Big[\mathbb{1}_{\{D_0\leq t-s\}}P(\{0\}, t-s)^{n-k}\Big] \\ &= \mathcal{E}_*\Big[P(\{0\}, t-s)^{n-k}\Big] \\ &= q_{0,n-k}(t-s)\end{aligned}$$

because $\mathbb{1}_{\{D_0>t-s\}}P(\{0\}, t-s)^{n-k} = 0$ almost-surely under $\mathcal{P}_*$.

The next step is to express $\mathbf{C}_{m_1,m_2,n}(t)$ in terms of the $\mathbf{C}_{j_1,j_2,k}$ functions, for each (j_1, j_2, k) satisfying $0 \leq j_1 \leq m_1, 0 \leq j_2 \leq m_2, 0 \leq k \leq n$, and $j_1 + j_2 + k < m_1 + m_2 + n$,. In order to do this rigorously we first assume that the distribution functions E_1, E_2, L_1, and L_2 are all continuous on $[0, \infty)$ (we will remove this assumption at the end of the proof, via a standard limiting argument). Differentiating with respect to t yields

$$\frac{\partial}{\partial t}\mathbf{C}_{m_1,m_2,n}(t) = (\mathbf{K}_0 + \mathbf{K}_1)\mathbf{C}_{m_1,m_2,n}(t) + \mathbf{F}(t)$$

where $\mathbf{F} : [0, \infty) \to |S| \times |S|$ is defined as

$$\begin{aligned}\mathbf{F}(t) := &\sum_{j_1=0}^{m_1-1}\binom{m_1}{j_1}\overline{E}_1(t)\mathbf{K}_1\mathbf{C}_{j_1,m_2,n}(t) + \sum_{j_2=0}^{m_2-1}\sum_{k=0}^{n}\binom{m_2}{j_2}\binom{n}{k}q_{1,n-k}(t)\mathbf{K}_1\mathbf{C}_{m_1,j_2,k}(t) \\ &+ \sum_{k=0}^{n-1}\binom{n}{k}q_{0,n-k}(t)\mathbf{K}_1\mathbf{C}_{m_1,m_2,k}(t).\end{aligned}$$

This matrix ODE can be solved using standard methods. Indeed, for each $t > 0$,

$$\mathbf{e}^{-(\mathbf{K}_0+\mathbf{K}_1)t}\frac{\partial}{\partial t}\mathbf{C}_{m_1,m_2,n}(t) - \mathbf{e}^{-(\mathbf{K}_0+\mathbf{K}_1)t}(\mathbf{K}_0 + \mathbf{K}_1)\mathbf{C}_{m_1,m_2,n}(t) = \mathbf{e}^{-(\mathbf{K}_0+\mathbf{K}_1)t}\mathbf{F}(t).$$

Since $\mathbf{e}^{\mathbf{A}t}\mathbf{A} = \mathbf{A}\mathbf{e}^{\mathbf{A}t}$ we have by the product rule and the fundamental theorem of calculus, that for each $t > 0$,

$$\mathbf{e}^{-(\mathbf{K}_0+\mathbf{K}_1)t}\mathbf{C}_{m_1,m_2,n}(t) - \mathbf{I}\mathbb{1}_{\{m_1=m_2=n=0\}} = \int_0^t \mathbf{e}^{-(\mathbf{K}_0+\mathbf{K}_1)s}\mathbf{F}(s)ds$$

i.e.

$$\mathbf{C}_{m_1,m_2,n}(t) = \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)t}\mathbb{1}_{\{m_1=m_2=n=0\}} + \int_0^t \mathbf{e}^{(\mathbf{K}_0+\mathbf{K}_1)(t-s)}\mathbf{F}(s)ds.$$

Replacing $\mathbf{F}(s)$ with the expression it represents completes the derivation. Finally, to handle the case where E_1, E_2, L_1, and L_2 are arbitrary, we can make use of the fact that each of these distribution functions can be interpreted as the weak limit of a sequence of phase-type distributions: see e.g. Chapter 3 of Kelly [35]. □

4 Models with Exponential mRNA lifetimes

In this section we further analyze the joint distribution of $I(t)$ and $M(t)$, when mRNA lifetimes are exponentially distributed with rate δ and elongation times are equal to zero. In this setting $\{(I(t), M(t)), t \geq 0\}$ is a continuous-time Markov chain.

For each real $t \geq 0$, each integer $m \geq 0$, and each $z \in \mathbb{D}$, define the matrices

$$\mathbf{J}(t,m) := \sum_{m_1=0}^{\infty}\sum_{n=0}^{\infty}\mathbf{J}(t,m_1,m,n) = \left[\mathbb{P}_x(M(t)=m, I(t)=y)\right]_{x,y\in S}$$

and

$$\hat{\mathbf{J}}(t,z) := \hat{\mathbf{J}}(t,1,z,1) = \left[\mathbb{E}_x[z^{M(t)}\mathbb{1}_{\{I(t)=y\}}]\right]_{x,y\in S}.$$

Instead of calculating the matrices $\mathbf{J}(t,m)$ and $\hat{\mathbf{J}}(t,z)$ directly, we focus on calculating their Laplace transforms. For each $\alpha \in \mathbb{C}_+ := \{\alpha \in \mathbb{C} : \mathrm{Re}(\alpha) > 0\}$ and each $z \in \mathbb{D}$, define the Laplace transform matrix $\mathbf{\Phi}(\alpha, z)$ as

$$\mathbf{\Phi}(\alpha,z) = [\Phi_{x,y}(\alpha)]_{x,y\in S} = \int_0^\infty e^{-\alpha t}\hat{\mathbf{J}}(t,z)\,dt,$$

where the integration is performed element-wise. This matrix is well defined for all $\alpha \in \mathbb{C}_+$. Next, for each integer $m \geq 0$, define the matrix $\mathbf{\Psi}_m(\alpha)$ (with integration performed element-wise) as

$$\mathbf{\Psi}_m(\alpha) := \int_0^\infty e^{-\alpha t}\mathbf{J}(t,m)\,dt.$$

This matrix is also well-defined for all $\alpha \in \mathbb{C}_+$, and clearly

$$\mathbf{\Phi}(\alpha,z) = \sum_{m=0}^{\infty} z^m \mathbf{\Psi}_m(\alpha).$$

Define the matrix $\mathbf{L}(\alpha)$ for each $\alpha \in \mathbb{C}_+$ as

$$\mathbf{L}(\alpha) := (\alpha\mathbf{I} - \mathbf{K}_0 - \mathbf{K}_1)^{-1}.$$

This matrix is well-defined for each $\alpha \in \mathbb{C}_+$, since $\mathbf{K}_0 + \mathbf{K}_1$ corresponds to the transition rate matrix of a finite-state CTMC. Readers should also recall that for each $x, y \in S$, the element found in row x, column y of $\mathbf{L}(\alpha)$, denoted $[\mathbf{L}(\alpha)]_{x,y}$, satisfies

$$[\mathbf{L}(\alpha)]_{x,y} = \int_0^\infty e^{-\alpha t}\mathbb{P}_x(I(t)=y)dt.$$

Given $j \in \mathbb{Z}_+$ and matrices $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{j-1}$, define $\prod_{i=0}^{j-1} \mathbf{A}_i$ to be the identity matrix if the product is empty (i.e. if $j = 0$) and otherwise

$$\prod_{i=0}^{n-1} \mathbf{A}_i = \mathbf{A}_0 \mathbf{A}_1 \ldots \mathbf{A}_{j-1}.$$

We are now ready to state the main result of this section.

Proposition 4.1. *For each $z \in \mathbb{D}$, and each $\alpha \in \mathbb{C}_+$,*

$$\mathbf{\Phi}(\alpha, z) = \sum_{j=0}^{\infty} (z-1)^j \left[\prod_{i=0}^{j-1} (\mathbf{L}(\alpha + i\delta)\mathbf{K}_1) \right] \mathbf{L}(\alpha + j\delta). \tag{18}$$

As a consequence, for each integer $m \geq 0$,

$$\mathbf{\Psi}_m(\alpha) = \sum_{j=0}^{\infty} (-1)^j \binom{j+m}{m} \left[\prod_{i=0}^{j+m-1} (\mathbf{L}(\alpha + i\delta)\mathbf{K}_1) \right] \mathbf{L}(\alpha + m + j\delta). \tag{19}$$

Proof. Firstly, note that when all elongation times are equal to zero, and the lifetime of each mRNA molecule is exponentially distributed with rate δ, we find that for each $t \geq 0$, and each $z \in \mathbb{C}_+$,

$$h(t, z) := \mathcal{E}_* \left[z^{\mathbb{1}\{D_0 \leq t, D_0 + B_0 > t\}} \right] = h(t, 1, z, 1) = 1 - e^{-\delta t} + z e^{-\delta t} = 1 - (1 - z) e^{-\delta t}$$

and plugging this expression into (6) yields

$$\frac{\partial}{\partial t} \hat{\mathbf{J}}(t, z) = \left(\mathbf{K}_0 + \mathbf{K}_1 - (1 - z) e^{-\delta t} \mathbf{K}_1 \right) \hat{\mathbf{J}}(t, z). \tag{20}$$

Equation (20) can be solved using Laplace transforms: after multiplying both sides of (20) by $e^{-\alpha t}$, then integrating with respect to t over $[0, \infty)$ (with respect to Lebesgue measure) while remembering the initial condition $\hat{\mathbf{J}}(0, z) = \mathbf{I}$, we get

$$\alpha \mathbf{\Phi}(\alpha, z) - \mathbf{I} = (\mathbf{K}_0 + \mathbf{K}_1)\, \mathbf{\Phi}(\alpha, z) - (1 - z) \mathbf{K}_1 \mathbf{\Phi}(\alpha + \delta, z). \tag{21}$$

Rearranging terms in (21) and multiplying both sides by $\mathbf{L}(\alpha)$ gives

$$\mathbf{\Phi}(\alpha, z) = \mathbf{L}(\alpha) - (1 - z) \mathbf{L}(\alpha) \mathbf{K}_1 \mathbf{\Phi}(\alpha + \delta, z). \tag{22}$$

Repeated applications of (22) yield, for each integer $n \geq 0$,

$$\begin{aligned} \mathbf{\Phi}(\alpha, z) = &\sum_{j=0}^{n} (z-1)^j \left[\prod_{\ell=0}^{j-1} (\mathbf{L}(\alpha + \ell\delta)\mathbf{K}_1) \right] \mathbf{L}(\alpha + j\delta) \\ &+ (z-1)^{n+1} \left[\prod_{\ell=0}^{n} (\mathbf{L}(\alpha + \ell\delta)\mathbf{K}_1) \right] \mathbf{\Phi}(\alpha + (n+1)\delta, z). \end{aligned} \tag{23}$$

In order to prove the first claim of the proposition, it suffices to show that as $n \to \infty$,

$$(z-1)^{n+1} \left[\prod_{\ell=0}^{n} (\mathbf{L}(\alpha + \ell\delta)\mathbf{K}_1) \right] \mathbf{\Phi}(\alpha + (n+1)\delta, z) \to \mathbf{0},$$

where the above convergence is element-wise. Throughout the argument, for a given square matrix $\mathbf{A}$, let $[\mathbf{A}]_{i,j}$ denote the element found in Row i, Column j of $\mathbf{A}$.

Set $C := \max_{i,j} |[\mathbf{K}_1]_{i,j}|$, and observe first that for each $\ell \geq 1$, and each $i, j \in \{0, 1, 2, \ldots, p\}$,

$$\begin{aligned}
|[\mathbf{L}(\alpha + \ell\delta)\mathbf{K}_1]_{i,j}| &= \left| \sum_{k=0}^{p} [\mathbf{L}(\alpha + \ell\delta)]_{i,k} [\mathbf{K}_1]_{k,j} \right| \leq \sum_{k=0}^{p} |[\mathbf{L}(\alpha + \ell\delta)]_{i,k}||[\mathbf{K}_1]_{k,j}| \\
&\leq C \sum_{k=0}^{p} |[\mathbf{L}(\alpha + \ell\delta)]_{i,k}| \leq C \sum_{k=0}^{p} \int_0^\infty e^{-(\mathrm{Re}(\alpha)+\ell\delta)t} \mathbb{P}_i(I(t) = k) dt \\
&= C \int_0^\infty e^{-(\mathrm{Re}(\alpha)+\ell\delta)} dt = \frac{C}{\mathrm{Re}(\alpha) + \ell\delta} < \frac{C}{\ell\delta}.
\end{aligned}$$

Having this observation in mind, observe next that for each $\ell_1, \ell_2 \geq 1$,

$$\begin{aligned}
|[\mathbf{L}(\alpha + \ell_1\delta)\mathbf{K}_1\mathbf{L}(\alpha + \ell_2\delta)\mathbf{K}_1]_{i,j}| &\leq \sum_{k=0}^{p} |[\mathbf{L}(\alpha + \ell_1\delta)\mathbf{K}_1]_{i,k}||[\mathbf{L}(\alpha + \ell_2\delta)\mathbf{K}_1]_{k,j}| \\
&\leq \frac{C^2}{\ell_1\ell_2\delta^2} \sum_{k=0}^{p} (1) < \frac{(p+1)C^2}{\ell_1\ell_2\delta^2}.
\end{aligned}$$

Using induction, it follows that for each $i, j \in \{0, 1, \ldots, p\}$,

$$\left[\prod_{\ell=1}^{n} (\mathbf{L}(\alpha + \ell\delta)\mathbf{K}_1) \right]_{i,j} \leq \frac{(p+1)^{n-1} C^n}{\delta^n n!} \to 0$$

as $n \to \infty$. Letting $n \to \infty$ in (23) gives

$$\mathbf{\Phi}(\alpha, z) = \sum_{j=0}^{\infty} (z-1)^j \left[\prod_{i=0}^{j-1} (\mathbf{L}(\alpha + i\delta)\mathbf{K}_1) \right] \mathbf{L}(\alpha + j\delta), \tag{24}$$

proving (18).

Finally, it follows from (18) that for each integer $m \geq 0$,

$$\mathbf{\Psi}_m(\alpha) = \frac{1}{m!} \frac{\partial^m}{\partial z^m} \mathbf{\Phi}(\alpha, z)\big|_{z=0} = \sum_{j=0}^{\infty} (-1)^j \binom{j+m}{m} \left[\prod_{i=0}^{j+m-1} (\mathbf{L}(\alpha + i\delta)\mathbf{K}_1) \right] \mathbf{L}(\alpha + m + j\delta) \tag{25}$$

which yields (19), thus proving Proposition 4.1. □

It is also easy to use Theorem 2.6 to derive a simple matrix recursion for the Laplace transform of the moments of mRNA molecules, when their elongation times are zero and their lifetimes are exponentially distributed with rate γ. For each integer $m \geq 0$, define

$$\boldsymbol{\mathcal{M}}_m(\alpha) := \int_0^\infty e^{-\alpha t} \mathbf{C}_{0,m,0}(t) dt.$$

Since (as in Theorem 2.6) $\mathbf{C}_{0,0,0}(t) = \mathbb{P}_x(I(t) = y) = e^{(\mathbf{K}_0+\mathbf{K}_1)t}$ we have

$$\boldsymbol{\mathcal{M}}_0(\alpha) = \int_0^\infty e^{-\alpha t} e^{(\mathbf{K}_0+\mathbf{K}_1)t} dt = \mathbf{L}(\alpha).$$

The following can be derived in a straightforward manner from Theorem 2.6.

Proposition 4.2. *For each integer $m \geq 0$,*

$$\boldsymbol{\mathcal{M}}_{m+1}(\alpha) = \sum_{j=0}^{m} \binom{m+1}{j} \mathbf{L}(\alpha)\mathbf{K}_1 \boldsymbol{\mathcal{M}}_j(\alpha+\delta).$$

We could have also used Proposition 4.1 to derive these (Laplace transforms of) moment matrices by taking derivatives with respect to z and setting $z = 1$, but this procedure would have given us factorial moment matrices instead of the moment matrices we actually wish to study. It is well-known that the moment matrices can be expressed in terms of the factorial moment matrices, but we suspect the recursion from Proposition 4.2 is easier to use.

Remark 4.1. We can use Proposition 4.1 to analyze the equilibrium joint distribution of the state of the gene and the mRNA level. Let $(I(\infty), M(\infty))$ denote a random vector whose law corresponds to the weak limit of the positive recurrent Markov chain $(I(t), M(t))$ as $t \to \infty$. Writing $\mathbf{J}(m) := \left[\mathbb{P}_x(M(\infty) = m, I(\infty) = y)\right]_{x,y\in S}$ and $\hat{\mathbf{J}}(z) := \hat{\mathbf{J}}(z,1) = \left[\mathbb{E}_x[z^{M(\infty)}\mathbb{1}_{\{I(\infty)=y\}}]\right]_{x,y\in S}$ for the stationary limits and applying (24) at $\alpha = \delta$ we obtain

$$\begin{aligned}\hat{\mathbf{J}}(z) &= \lim_{\alpha\to 0} \alpha\boldsymbol{\Phi}(\alpha) = \lim_{\alpha\to 0} \alpha\mathbf{L}(\alpha)\cdot(\mathbf{I} - (1-z)\mathbf{K}_1\boldsymbol{\Phi}(\delta)) \\ &= \hat{\mathbf{J}}(1)\left(\mathbf{I} + \sum_{j=1}^{\infty}(z-1)^j\mathbf{K}_1\left[\prod_{i=1}^{j-1}(\mathbf{L}(i\delta)\mathbf{K}_1)\right]\mathbf{L}(j\delta)\right). \end{aligned} \tag{26}$$

where $\hat{\mathbf{J}}(1) = \lim_{\alpha\to 0}\alpha\mathbf{L}(\alpha) = [\mathbb{P}_x(I(\infty) = y)]_{x,y\in S}$ is the matrix consisting of n equal rows representing the stationary distribution of I.

Formula (18) simplifies considerably when $\mathbf{K}_1$ is of rank one (i.e. when there exist two column vectors $\mathbf{v}$, $\mathbf{w}$ satisfying $\mathbf{K}_1 = \mathbf{v}\mathbf{w}^\intercal$). In many cases this simplification leads to expressions in terms of generalized hypergeometric functions.

Definition 4.1. *The generalized hypergeometric function ${}_mF_n$ associated with the complex numbers $a_1, a_2, \ldots, a_m$, $b_1, b_2, \ldots, b_n$, is a power series that is defined as follows*:

$${}_mF_n(a_1,\ldots,a_m;b_1,\ldots,b_n;z) := \sum_{k=0}^{\infty}\left[\frac{\prod_{\ell=1}^{m}(a_\ell)^{(k)}}{\prod_{\ell'=1}^{n}(b_{\ell'})^{(k)}}\right]\frac{z^k}{k!}$$

where for each $x \in \mathbb{C}$, $(x)^{(0)} := 1$, and $(x)^{(k)} := x(x+1)\cdots(x+(k-1))$ for each integer $k \geq 1$.

Generalized hypergeometric functions are defined and discussed in Graham et al. [17].

Theorem 4.2. *Suppose $\mathbf{K}_1 \in \mathbb{R}^{p+1\times p+1}$ is of the form $\mathbf{K}_1 = \mathbf{v}\mathbf{w}^\intercal$, where $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{p+1}$. Then*

$$\boldsymbol{\Phi}(\alpha, z) = \mathbf{L}(\alpha) + \mathbf{L}(\alpha)\mathbf{v}\sum_{j=1}^{\infty}(z-1)^j b_j(\alpha)\mathbf{w}^\intercal\mathbf{L}(\alpha+j\delta), \tag{27}$$

where $b_j(\alpha) = \prod_{i=1}^{j-1}\operatorname{tr}(\mathbf{L}(\alpha+i\delta)\mathbf{K}_1)$ and $\operatorname{tr}(\cdot)$ denotes the trace function.

Proof. Let $\mathbf{B} \in \mathbb{R}^{p+1\times p+1}$ be arbitrary. Writing $\mathbf{w} = (w_0,\ldots,w_p)^\intercal$, $\mathbf{v} = (v_0,\ldots,v_p)^\intercal$, it is clear that the element found in Row i, Column j of $\mathbf{v}\mathbf{w}^\intercal$ is $v_i w_j$. Since

$$\operatorname{tr}(\mathbf{B}\mathbf{K}_1) = \sum_{i=0}^{p}\sum_{\ell=0}^{p} b_{i,\ell}v_\ell w_i = \sum_{i=0}^{p} w_i \sum_{\ell=0}^{p} b_{i,\ell}v_\ell = \mathbf{w}^\intercal\mathbf{B}\mathbf{v}$$

we get

$$\mathbf{K}_1\mathbf{B}\mathbf{K}_1 = \mathbf{v}(\mathbf{w}^\top\mathbf{B}\mathbf{v})\mathbf{w}^\top = (\mathbf{w}^\top\mathbf{B}\mathbf{v})\mathbf{v}\mathbf{w}^\top = (\mathbf{w}^\top\mathbf{B}\mathbf{v})\mathbf{K}_1 = \mathrm{tr}(\mathbf{B}\mathbf{K}_1)\mathbf{K}_1$$

which in turn implies, for each $j \geq 1$,

$$\prod_{i=0}^{j-1}(\mathbf{L}(\alpha+i\delta)\mathbf{K}_1) = \left(\prod_{i=1}^{j-1}\mathrm{tr}(\mathbf{L}(\alpha+i\delta)\mathbf{K}_1)\right)\mathbf{L}(\alpha)\mathbf{K}_1.$$

Formula (27) then follows from (24). □

4.1 A Two-State Environment

In our next corollary, we study the joint distribution of the state of the gene and the number of mRNA molecules present at time t, for the case where the matrix $\mathbf{K}_1 = \mathbf{v}\mathbf{w}^\top$. This model is a generalization of the telegraph model of Peccoud and Ycart [41]. Note that the time-dependent behavior of the telegraph model was addressed in Iyer-Biswas et al. [25].

Corollary 4.1. *Suppose that*

$$\mathbf{K}_0 = \begin{pmatrix} -\lambda - c(\rho+\tau) & \lambda \\ \mu & -\mu - d(\rho+\tau) \end{pmatrix}, \quad \mathbf{K}_1 = \begin{pmatrix} c\rho & c\tau \\ d\rho & d\tau \end{pmatrix},$$

for $\lambda, \mu, c, d, \rho, \tau \in [0,\infty)$. *Then*

$$\begin{aligned}\mathbf{\Phi}(\alpha,z) = {} & \frac{1}{\alpha(\alpha+\lambda+\mu+d\rho+c\tau)}\left[\begin{pmatrix} \alpha+\mu+d\rho & \lambda+c\tau \\ \mu+d\rho & \alpha+\lambda+c\tau \end{pmatrix} + \right. \\ & \times \frac{1}{\alpha(c\rho+d\tau)+(\rho+\tau)(c\mu+d(\lambda+c(\rho+\tau))} \\ & \times \sum_{j=1}^{\infty}\left((z-1)^j \frac{\prod_{i=0}^{j-1}(\alpha(c\rho+d\tau)+i\delta(c\rho+d\tau)+(\rho+\tau)(c\mu+d(\lambda+c(\rho+\tau)))}{\prod_{i=0}^{j-1}(\alpha+\delta+i\delta)\prod_{i=0}^{j-1}(\alpha+\delta+\lambda+\mu+d\rho+c\tau+i\delta)}\right. \\ & \left.\left. \times \begin{pmatrix} d\lambda + c(\alpha+\mu+d(\rho+\tau)) \\ c\mu + d(\alpha+\lambda+c(\rho+\tau)) \end{pmatrix}\begin{pmatrix} (\alpha+j\delta)\rho + (\mu+d\rho)(\rho+\tau) \\ (\alpha+j\delta)\tau + (\lambda+c\tau)(\rho+\tau) \end{pmatrix}^\top\right)\right]. \end{aligned} \tag{28}$$

Proof. We have the situation of Theorem 4.2 with $\mathbf{K}_1 = \mathbf{v}\mathbf{w}^\top$, $\mathbf{v} = (c,d)^\top$ and $\mathbf{w}^\top = (\rho,\tau)$. In this case

$$\mathbf{L}(\alpha) = \frac{1}{C(\alpha)}\begin{pmatrix} \alpha+\mu+d\rho & \lambda+c\tau \\ \mu+d\rho & \alpha+\lambda+c\tau \end{pmatrix}$$

where $C(\alpha) = \alpha(\alpha+\lambda+\mu+d\rho+c\tau)$. Since

$$\mathbf{L}(\alpha)\mathbf{K}_1 = \frac{1}{C(\alpha)}\begin{pmatrix} \rho(d\lambda+c(\alpha+\mu+d(\rho+\tau))) & \tau(d\lambda+c(\alpha+\mu+d(\rho+\tau))) \\ \rho(c\mu+d(\alpha+\lambda+c(\rho+\tau))) & \tau(c\mu+d(\alpha+\lambda+c(\rho+\tau))) \end{pmatrix}$$

it follows that

$$\mathrm{tr}(\mathbf{L}(\alpha)\mathbf{K}_1) = \frac{\alpha(c\rho+d\tau)+(\rho+\tau)(c\mu+d(\lambda+c(\rho+\tau))}{\alpha(\alpha+\lambda+\mu+d\rho+c\tau)}.$$

Moreover

$$\mathbf{L}(\alpha)\mathbf{v} = \frac{1}{C(\alpha)}\begin{pmatrix} d\lambda + c(\alpha+\mu+d(\rho+\tau)) \\ c\mu + d(\alpha+\lambda+c(\rho+\tau)) \end{pmatrix}, \quad \mathbf{w}^\intercal\mathbf{L}(\alpha) = \frac{1}{C(\alpha)}\begin{pmatrix} \alpha\rho + (\mu+d\rho)(\rho+\tau) \\ \alpha\tau + (\rho+\tau)(\lambda+c\tau) \end{pmatrix}^\intercal.$$

Then (28) follows from (27) after some simplifications. □

Example 4.3 (Extended Telegraph Model)**.** Consider the following extension of the telegraph model. The gene switches from inactive to active with rate $\lambda+\lambda_*$. Here λ is associated with simple transitions of $I(t)$ whereas λ_* is the rate at which a transition to the active state happens together with a simultaneous production of an mRNA. As in the telegraph model (where $\lambda_* = 0$), in the active state mRNAs are produced with rate K_A. We thus obtain the matrices

$$\mathbf{K}_0 = \begin{pmatrix} -(\lambda+\lambda_*) & \lambda \\ \mu & -(\mu+K_A) \end{pmatrix}, \quad \mathbf{K}_1 = \begin{pmatrix} 0 & \lambda_* \\ 0 & K_A \end{pmatrix}. \tag{29}$$

In terms of the above corollary $c = \lambda_*/K_A$, $d = 1$, $\rho = 0$ and $\tau = K_A$, so that after some calculations we obtain from (28)

$$\begin{aligned}\mathbf{\Phi}(\alpha,z) = {} & \frac{1}{\alpha+\beta}\begin{pmatrix} 1 & -c \\ 0 & 0 \end{pmatrix} + \frac{1}{\alpha(\alpha+\mu+\lambda+\lambda_*)} \\ & \times \sum_{j=0}^{\infty}\left(\frac{(K_A(z-1))^j\prod_{i=0}^{j-1}(\alpha+\beta+i\delta)}{\prod_{i=1}^{j}(\alpha+i\delta)\prod_{i=1}^{j}(\alpha+\mu+\lambda+\lambda_*+i\delta)}\begin{pmatrix} \frac{c\alpha+\beta}{\alpha+\beta} \\ 1 \end{pmatrix}\begin{pmatrix} \mu \\ \alpha+j\delta+\lambda+\lambda_* \end{pmatrix}^\intercal\right)\end{aligned}$$

where we used the abbreviation $\beta = \lambda + \lambda_* + \lambda_*\mu/K_A$. We can express the components of the matrix $\mathbf{\Phi}(\alpha,z)$ in terms of generalized hypergeometric functions. Letting

$$\begin{aligned} H_0(\alpha,z) &= \frac{1}{\alpha(\alpha+\lambda+\mu+\lambda_*)}\,{}_2F_2\left(1, \tfrac{\alpha+\beta}{\delta}; 1+\tfrac{\alpha}{\delta}, 1+\tfrac{\alpha+\lambda+\lambda_*+\mu}{\delta}; \tfrac{K_A(z-1)}{\delta}\right), \\ H_1(\alpha,z) &= \frac{1}{\alpha(\alpha+\lambda+\mu+\lambda_*)}\,{}_2F_2\left(1, 1+\tfrac{\alpha+\beta}{\delta}; 1+\tfrac{\alpha}{\delta}, 1+\tfrac{\alpha+\lambda+\lambda_*+\mu}{\delta}; \tfrac{K_A(z-1)}{\delta}\right), \end{aligned}$$

the entries $\Phi_{x,y}$ of $\mathbf{\Phi}$ are linear combinations of H_0 and H_1:

$$\begin{aligned} \Phi_{0,0}(\alpha,z) &= \frac{1}{\alpha+\beta} + \frac{\mu(c\alpha+\beta)}{\alpha+\beta}H_0(\alpha,z), \\ \Phi_{0,1}(\alpha,z) &= -\frac{c}{\alpha+\beta} + (c\alpha+\beta)\,H_1(\alpha,z) - \mu c\frac{c\alpha+\beta}{\alpha+\beta}H_0(\alpha,z), \\ \Phi_{1,0}(\alpha,z) &= \mu H_0(\alpha,z), \\ \Phi_{1,1}(\alpha,z) &= (\alpha+\beta)H_1(\alpha,z) - \mu c H_0(\alpha,z). \end{aligned}$$

Since $\frac{\partial}{\partial z}\,{}_2F_2(a_1,a_2;b_1,b_2;z) = \frac{a_1a_2}{b_1b_2}\,{}_2F_2(a_1+1,a_2+1;b_1+1,b_2+1;z)$, higher derivatives of the generalized hypergeometric function are given by

$$\frac{\partial^n}{\partial z^n}\,{}_2F_2\left(a_1,a_2;b_1,b_2;z\right) = \frac{(a_1)_n(a_2)_n}{(b_1)_n(b_2)_n}\,{}_2F_2\left(a_1+n,a_2+n;b_1+n,b_2+n;z\right).$$

With this the calculation of the transient probability transform matrix $\mathbf{\Psi}_n(\alpha)$ is straightforward.

Letting $\alpha \to 0$ in $\alpha\mathbf{\Phi}(\alpha, z)$ we obtain the stationary probabilities in terms of confluent hypergeometric functions of the first kind,

$$\mathbb{E}[z^{M(\infty)}\mathbb{1}_{\{I(\infty)=0\}}] = \frac{\mu}{\lambda+\mu+\lambda_*}\,{}_1F_1\left(\tfrac{\beta}{\delta}; 1+\tfrac{\lambda+\lambda_*+\mu}{\delta}; \tfrac{K_A(z-1)}{\delta}\right),$$
$$\mathbb{E}[z^{M(\infty)}\mathbb{1}_{\{I(\infty)=1\}}] = \frac{1}{\lambda+\mu+\lambda_*}\left(\beta\,{}_1F_1\left(1+\tfrac{\beta}{\delta}; 1+\tfrac{\lambda+\lambda_*+\mu}{\delta}; \tfrac{K_A(z-1)}{\delta}\right)\right.$$
$$\left. - \mu c\,{}_1F_1\left(\tfrac{\beta}{\delta}; 1+\tfrac{\lambda+\lambda_*+\mu}{\delta}; \tfrac{K_A(z-1)}{\delta}\right)\right).$$

In particular, as is obvious from the setting anyway, $\mathbb{P}(I(\infty)=0) = \mu/(\mu+\lambda+\lambda_*)$ and $\mathbb{P}(I(\infty)=0) = (\lambda+\lambda_*)/(\mu+\lambda+\lambda_*)$.

4.2 A Three-State Example

We now revisit Example 2.2, which corresponds to the model analyzed in Cao et al. [10]. In our context, the current state of the gene and the creation of mRNA molecules are governed by a MAP with matrices

$$\mathbf{K}_0 = \begin{pmatrix} -a_1 & a_1 & 0 \\ a_0 & -(a_0+a_2) & a_2 \\ a_0 & 0 & -(a_0+\rho) \end{pmatrix}, \quad \mathbf{K}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \rho & 0 \end{pmatrix}.$$

Clearly $\mathbf{K}_1$ satisfies $\mathbf{K}_1 = \mathbf{v}\mathbf{w}^\tau$ where $\mathbf{v} = (0,0,1)^\tau$ and $\mathbf{w}^\tau = (0,\rho,0)$. Letting $C(\alpha) = \alpha(a_0+a_1+\alpha)(a_0+a_2+\alpha+\rho)$, we obtain

$$\mathbf{L}(\alpha) = \frac{1}{C(\alpha)}\begin{pmatrix} (a_0+\alpha)(a_0+a_2+\alpha+\rho) & a_1(a_0+\alpha+\rho) & a_1a_2 \\ a_0(a_0+a_2+\alpha+\rho) & (a_1+\alpha)(a_0+\alpha+\rho) & a_2(a_1+\alpha) \\ a_0(a_0+a_2+\alpha+\rho) & a_0a_1+(a_1+\alpha)\rho & a_0\alpha+(a_1+\alpha)(a_2+\alpha) \end{pmatrix}$$
$$\mathbf{L}(\alpha)\mathbf{v} = \frac{1}{C(\alpha)}\begin{pmatrix} a_1a_2 \\ a_2(a_1+\alpha) \\ a_0\alpha+(a_1+\alpha)(a_2+\alpha) \end{pmatrix}$$
$$\mathbf{w}^\tau\mathbf{L}(\alpha) = \frac{\rho}{C(\alpha)}\begin{pmatrix} a_0(a_0+a_2+\alpha+\rho), (a_1+\alpha)(a_0+\alpha+\rho), a_2(a_1+\alpha) \end{pmatrix}.$$

Since $\mathrm{tr}(\mathbf{L}(\alpha)\mathbf{K}_1) = \frac{a_2(a_1+\alpha)\rho}{C(\alpha)}$ it follows that

$$b_j(\alpha) = \prod_{i=1}^{j-1}\mathrm{tr}(\mathbf{L}(\alpha+i\delta)\mathbf{K}_1) = (\rho a_2)^{j-1}\prod_{i=1}^{j-1}\frac{a_1+\alpha+i\delta}{C(\alpha+i\delta)}.$$

By (27) we then obtain

$$\mathbf{\Phi}(\alpha,z) = \mathbf{L}(\alpha) + \frac{1}{(a_1+\alpha)C(\alpha)}\begin{pmatrix} a_1a_2 \\ a_2(a_1+\alpha) \\ a_0\alpha+(a_1+\alpha)(a_2+\alpha) \end{pmatrix}$$
$$\times\sum_{j=1}^{\infty}\frac{(\rho a_2(z-1))^j\prod_{i=0}^{j-1}(a_1+\alpha+i\delta)}{\prod_{i=0}^{j-1}C(\alpha+\delta+i\delta)}\begin{pmatrix} a_0(a_0+a_2+\alpha+j\delta+\rho) \\ (a_1+\alpha+j\delta)(a_0+\alpha+j\delta+\rho) \\ a_2(a_1+\alpha+j\delta) \end{pmatrix}^\tau.$$

From this expression one can calculate the elements of $\mathbf{\Phi}$ and express them in terms of hypergeometric functions. For example $\Phi_{0,2}$ becomes

$$\begin{aligned}\Phi_{0,2}(\alpha,z) &= \frac{a_1a_2}{C(\alpha)} + \frac{a_1a_2}{(a_1+\alpha)C(\alpha)}\sum_{j=1}^{\infty}\frac{(\rho a_2(z-1))^j\prod_{i=0}^{j-1}(a_1+\alpha+i\delta)}{\prod_{i=0}^{j-1}C(\alpha+\delta+i\delta)}(a_1+\alpha+j\delta)\\ &= \frac{a_1a_2}{C(\alpha)}\left(1+\sum_{j=1}^{\infty}\frac{(\rho a_2(z-1))^j}{j!}\frac{\prod_{i=0}^{j-1}(1+i)\prod_{i=0}^{j-1}(1+\frac{a_1+\alpha}{\delta}+i)}{\prod_{i=0}^{j-1}C(\alpha+\delta+i\delta)}\right)\\ &= \frac{a_1a_2}{C(\alpha)}{}_2F_3\left(1,1+\frac{a_1+\alpha}{\delta};1+\frac{\alpha}{\delta},1+\frac{\alpha+a_0+a_1}{\delta},1+\frac{\alpha+a_0+a_2+\rho}{\delta},\frac{\rho a_2(z-1)}{\delta^2}\right).\end{aligned}$$

4.3 The Off/On/Seq-L model of De Gunst et al.

Recently, in De Gunst et al. [18] the authors introduce an interesting modification of the telegraph model. In their model, the gene alternates being active and inactive, but each time an mRNA molecule is conceived, it is a nascent mRNA molecule for an amount of time that is hypoexponentially distributed (the distribution of a sum of independent exponentially distributed random variables, where the rates among the sum may be distinct). Furthermore, in their model it is impossible for there to be more than one nascent mRNA molecule present in the system at any time, which basically means no other mRNA molecules can be conceived while a nascent mRNA molecule is present. Finally, as soon as the nascent mRNA molecule becomes a mature mRNA molecule, the gene resumes operating in the active state. The authors of [18] approach the problem by observing that if you keep track of both the state of the gene, and the number of mature mRNA molecules present in the cell at time t, you can model this system as a level-dependent Quasi-Birth-Death (QBD) process.

The Off/On/Seq-L model falls within our framework if we set all elongation times in our model equal to zero, while simultaneously incorporating the hypoexponential elongation times into the state space of the gene (i.e. the MAP). This is possible because in this model, no more than one nascent mRNA molecule can exist. All mRNA lifetimes are exponentially distributed with rate μ, independently of everything else.

If we put the On/Off/Seq-L model in our context, we let the state space of the MAP be $S = \{0,1,2,,\ldots,L\}$, and define

$$\mathbf{K}_0 = \begin{pmatrix} -q_0 & q_0 & 0 & 0 & 0 & \cdots & 0\\ q_1 & -(q_1+\lambda_1) & \lambda_1 & 0 & 0 & 0 & 0\\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & 0 & 0\\ 0 & 0 & 0 & -\lambda_3 & \lambda_3 & \ddots & 0\\ 0 & 0 & 0 & 0 & -\lambda_4 & \ddots & 0\\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots\\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_L \end{pmatrix}, \quad \mathbf{K}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0\\ 0 & \lambda_L & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Again, $\mathbf{K}_1$ is a matrix having rank one and we can express $\mathbf{K}_1$ as $\mathbf{K}_1 = \mathbf{v}\mathbf{w}^\intercal$, where

$$\mathbf{v} = (0,0,\ldots,1)^\intercal, \quad \mathbf{w} = (0,\lambda_L,0,\ldots,0)^\intercal.$$

It follows that $\mathbf{L}(\alpha)\mathbf{v} = ((\mathbf{L}(\alpha))_{0,L},\ldots,(\mathbf{L}(\alpha))_{L,L})^\intercal$ and $\mathbf{w}^\intercal\mathbf{L}(\alpha) = \lambda_L((\mathbf{L}(\alpha))_{1,0},\ldots,(\mathbf{L}(\alpha))_{1,L})$. Furthermore, $\mathrm{tr}(\mathbf{L}(\alpha)\mathbf{K}_1) = \lambda_L(\mathbf{L}(\alpha))_{1,L}$. The Laplace transform matrix $\mathbf{L}(\alpha)$ is very tractable for

this model. Indeed,

$$[\mathbf{L}(\alpha)]_{1,1} = \left[\alpha\left(1+\frac{q_1}{q_0+\alpha}\right)+\lambda_1\left(1-\prod_{\ell=2}^{L}\frac{\lambda_\ell}{\lambda_\ell+\alpha}\right)\right]^{-1}, \qquad [\mathbf{L}(\alpha)]_{1,0} = \frac{q_1}{q_0+\alpha}[\mathbf{L}(\alpha)]_{1,1}$$

and for $2 \le y \le L$,

$$[\mathbf{L}(\alpha)]_{1,y} = \frac{\lambda_1}{\lambda_y+\alpha}\prod_{\ell=2}^{y-1}\frac{\lambda_\ell}{\lambda_\ell+\alpha}[\mathbf{L}(\alpha)]_{1,1}.$$

Finally,

$$[\mathbf{L}(\alpha)]_{0,L} = \frac{q_0}{q_0+\alpha}\frac{\lambda_1}{\lambda_L+\alpha}\prod_{\ell=2}^{L-1}\frac{\lambda_\ell}{\lambda_\ell+\alpha}[\mathbf{L}(\alpha)]_{1,1}$$

and for $2 \le x \le L$,

$$[\mathbf{L}(\alpha)]_{x,L} = \frac{\lambda_1+\alpha\left(1+\frac{q_1}{q_0+\alpha}\right)}{\lambda_L+\alpha}\prod_{\ell=x}^{L-1}\frac{\lambda_\ell}{\lambda_\ell+\alpha}[\mathbf{L}(\alpha)]_{1,1}.$$

These formulas can be used to write down a series representation for each element of $\mathbf{\Phi}(\alpha, z)$:

$$\begin{aligned}\mathbf{\Phi}(\alpha,z) &= \sum_{j=0}^{\infty}(z-1)^j\left(\prod_{i=0}^{j-1}\mathbf{L}(\alpha+i\delta)\mathbf{K}_1\right)\mathbf{L}(\alpha+j\delta)\\ &= \mathbf{L}(\alpha)+\sum_{j=1}^{\infty}(z-1)^j\left(\prod_{i=1}^{j-1}\mathrm{tr}(\mathbf{L}(\alpha+i\delta)\mathbf{K}_1)\right)\mathbf{L}(\alpha)\mathbf{K}_1\mathbf{L}(\alpha+j\delta)\end{aligned}$$

which implies

$$[\mathbf{\Phi}(\alpha,z)]_{x,y} = [\mathbf{L}(\alpha)]_{x,y}+[\mathbf{L}(\alpha)]_{x,L}\sum_{j=1}^{\infty}(z-1)^j\lambda_L^j\left(\prod_{i=1}^{j-1}[\mathbf{L}(\alpha+i\delta)]_{1,L}\right)[\mathbf{L}(\alpha+j\delta)]_{1,y}.$$

Hence,

$$[\mathbf{\Psi}_0(\alpha)]_{x,y} = [\mathbf{L}(\alpha)]_{x,y}+[\mathbf{L}(\alpha)]_{x,L}\sum_{j=1}^{\infty}(-1)^j\lambda_L^j\left(\prod_{i=1}^{j-1}[\mathbf{L}(\alpha+i\delta)]_{1,L}\right)[\mathbf{L}(\alpha+j\delta)]_{1,y}$$

and for each integer $m \ge 1$,

$$[\mathbf{\Psi}_m(\alpha)]_{x,y} = [\mathbf{L}(\alpha)]_{x,L}\sum_{k=0}^{\infty}\binom{k+m}{m}(-1)^k\lambda_L^{k+m}\left(\prod_{i=1}^{k+m-1}[\mathbf{L}(\alpha+i\delta)]_{1,L}\right)[\mathbf{L}(\alpha+(k+m)\delta)]_{1,y}.$$

Multiplying both sides by α, then letting $\alpha \downarrow 0$ yields

$$\lim_{t\to\infty}\mathbb{P}_x(I(t)=y, M(t)=0) = p(y)+p(L)\sum_{j=1}^{\infty}(-1)^j\lambda_L^j\left(\prod_{i=1}^{j-1}[\mathbf{L}(i\delta)]_{1,L}\right)[\mathbf{L}(j\delta)]_{1,y} \qquad (30)$$

and for each integer $m \geq 1$,

$$\lim_{t\to\infty} \mathbb{P}_x(I(t) = y, M(t) = m) = p(L) \sum_{k=0}^{\infty} \binom{k+m}{m} (-1)^k \lambda_L^{k+m} \left(\prod_{i=1}^{k+m-1} [\mathbf{L}(i\delta)]_{1,L} \right) [\mathbf{L}((k+m)\delta)]_{1,y}. \tag{31}$$

In these expressions $p(y) = \mathbb{P}(I(\infty) = y)$ denotes the stationary distribution (in particular this does not depend on x) for the process $\{I(t); t \geq 0\}$, which is given by

$$p(1) = \frac{q_0}{q_1 + q_0 \sum_{y=1}^{L} \frac{\lambda_1}{\lambda_y}}, \qquad p(y) = \begin{cases} \frac{q_1}{q_0} p(1) & ; y = 0 \\ \frac{\lambda_1}{\lambda_y} p(1) & ; y \in \{2, 3, \ldots, L\} \end{cases}. \tag{32}$$

4.4 Simulation and numerical results for the Off/On/Seq-L model

In this section we chose the parameters for the Off/On/Seq-L model to be

$$q_0 = 0.7, q_1 = 1.1, \lambda_1 = 1.2, \lambda_2 = 1.6, \lambda_3 = 3.5, \gamma_1 = 0.1 \tag{33}$$

unless otherwise indicated. We chose these parameters in an admittedly arbitrary manner, in order to generate an interesting distribution of the number of mRNA molecules both at each finite time, and in equilibrium. We also assume throughout that $I(0) = 0$.

Using (33) we calculated the stationary probabilities $\mathbb{P}(M(\infty) = m)$ and compared them with simulation runs. As shown in Fig.1,(a) and (b), the numerical results coincide well with simulated values. For the parameters chosen a Poisson distribution with mean $\mathbb{E}[M(\infty)]$ also fits quite well. Figure (c) shows the simulated values for $\mathbb{P}_0(M(t) = m)$ for different values of t.

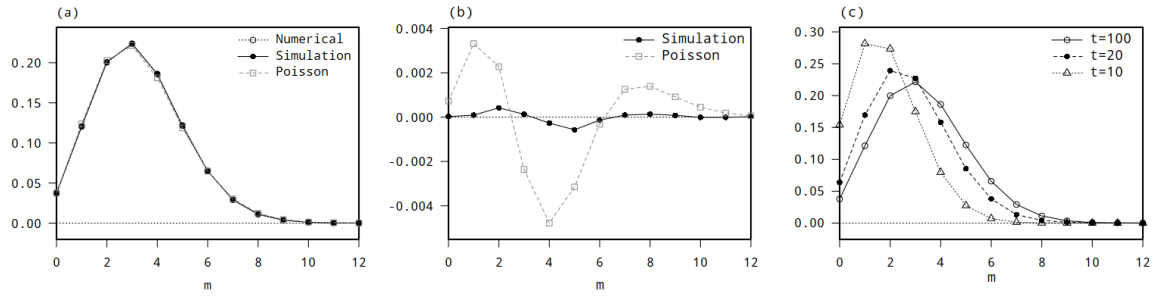

Fig. 1: (a) *The stationary probabilities* $\mathbb{P}(M(\infty) = m)$ *for* $m = 0, 1, 2, \ldots, 12$*, calculated numerically from* (30) *and* (31) *and derived from a simulation with one million samples (with simulation stopped at time* $t = 500$*). Additionally the same probabilities are calculated for a Poisson distribution with mean* $\mathbb{E}[M(\infty)] \approx 3.275$*.* **(b)** *as before but showing the difference w.r.t. the numerical solution.* **(c)** *Probabilities* $\mathbb{P}_x(M(t) = m)$ *for* $m = 0, 1, 2, \ldots, 12$ *at different times t, obtained from a simulation with 100 000 samples.*

Figure 2 compares the stationary distribution for the number of mRNAs for different mRNA lifetime distributions. We performed a simulation with time horizon $t = 100$ and 100 000 samples with exponential lifetimes, Bernoulli lifetimes and deterministic lifetimes (all with the same mean). For the parameters chosen the three lifetime distributions lead to almost identical limit distributions for $M(t)$.

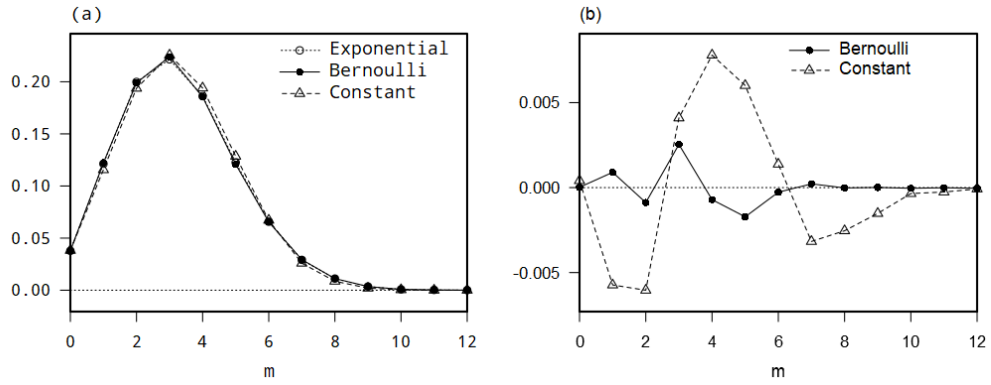

Fig. 2: **(a)** *Simulated distribution of* $M(t)$*,* $t = 100$ *with 100 000 samples. Different lifetime distributions for the mRNAs with mean 10: exponential distribution with parameter* $\gamma_1 = 0.1$*, Bernoulli distribution with* $\mathbb{P}(B^{(1)} = 1) = \mathbb{P}(B^{(1)} = 19) = 0.5$*, deterministic lifetime:* $\mathbb{P}(B^{(1)} = 10) = 1$*.* **(b)** *as in* **(a)** *but now differences w.r.t. the exponential case.*

To show the sensitivity of the limiting distribution of the number of mRNA to changes in parameters, we varied one parameter with the others held fixed (with values as in (33)). Figure 3 shows how the probabilities $\mathbb{P}(M(\infty) = m)$, the mean $\mathbb{E}[M(\infty)]$ and the variance $\mathrm{Var}(M(\infty))$ differ for varying q_0, q_1 and γ_1.

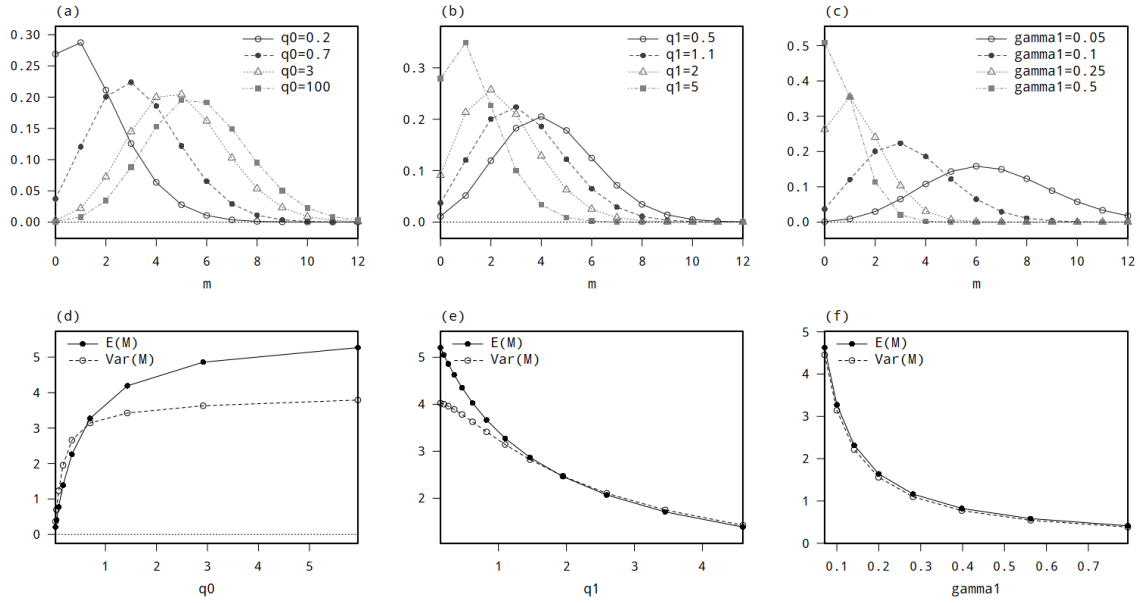

Fig. 3: *Top row: Stationary probabilities* $\mathbb{P}(M(\infty) = m)$ *for* $m = 0, 1, 2, \ldots, 12$ *as obtained from* (30) *and* (31)*. Bottom row: Expected values and variances for* $M(\infty)$ *calculated from the probabilities. The parameter* q_0 *is varied in* **(a)**, **(d)***;* q_1 *in* **(b)**, **(e)***;* γ_1 *in* **(c)**, **(f)**.

We observe from Figure 3 (d) that as q_0 increases, while the other values within (33) remain fixed, the law of $M(\infty)$ becomes 'less-and-less' Poisson. If we now reset the value of q_0 to be $q_0 = 0.7$, the distribution of $M(\infty)$ appears to be almost Poisson. In the interest of seeing what happens to the law of $M(t)$ for each finite t, we used Theorem 2.6 to approximate the mean, variance, and Fano factor

(ratio of variance to mean) of $M(t)$. The resulting curves, sketched over the set $0 < t < 100$, are given below: these were generated while assuming $I(0) = 0$ with probability one.

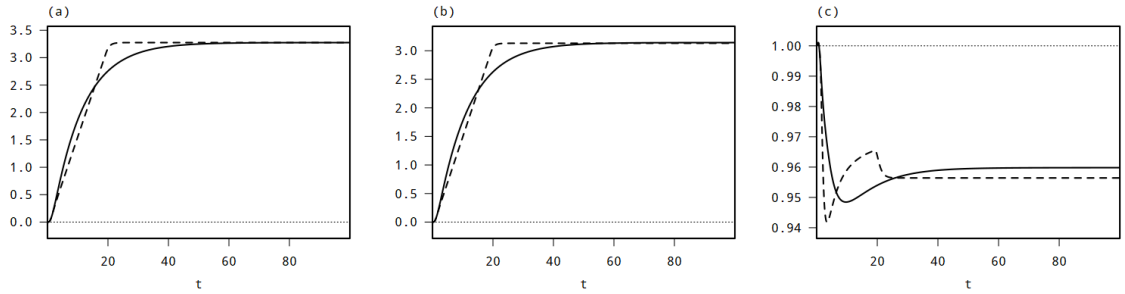

Fig. 4: *Time dependent characteristics of the mRNA number for exponential lifetimes (solid) and Bernoulli lifetime (dashed)* **(a)** $\mathbb{E}[M(t)]$ **(b)** $Var(M(t))$ **(c)** $Var(M(t))/\mathbb{E}[M(t)]$

It is clear that for both types of mRNA lifetime distributions, the Fano factor of $M(t)$ is consistently about 0.04 units below unity for all but very small values of t.

5 Conclusion

We have shown that, for a general three-stage model of stochastic gene expression, various quantities can be derived that shed light on the time-dependent joint distribution of the current state of the gene, the number of mature mRNA molecules present in the cell, and the number of mature protein molecules present in the cell. Our main results establish interesting formulas when mRNA molecules are conceived in accordance with an arbitrary Markovian Arrival Process, where the elongation times and lifetimes of the mRNA molecules are generally distributed, and when the elongation times and lifetimes of the protein molecules are generally distributed. Markovian Arrival Processes are used in order to construct models where mRNA molecules can be conceived both at various instants at which the gene makes a state transition, as well as in a Poisson manner where the rate depends on the current state of the gene.

We then showed that when we only track mRNA molecules and assume negligible elongation times and exponentially distributed lifetimes, even more elegant expressions can be derived for the time-dependent joint distributions, and the joint stationary distribution, of the state of the gene and the number of mRNA molecules present in the cell. It is notable that many recent gene expression models fall within our framework, and we hope that the ideas found herein will inspire others to build more elaborate models that capture dynamics our models do not currently capture. For example, it is not clear how to use the analysis provided here to study models where the state of the gene is further influenced by the number of proteins currently found in the cell: simpler Markovian models that account for such feedback mechanisms include (but are not limited to) the recent works [29, 31]. It would also be interesting to see to what extent recent studies of models that incorporate cell division, such as [5, 42], can be generalized while simultaneously providing new insights into gene expression.

Data Availability Statement

No data was collected while performing this research.

Acknowledgements

The authors thank 3 anonymous referees for helpful comments that improved the paper. MH was supported in part by Future Fellowship FT160100166 from the Australian Research Council.

References

[1] Adan, I., and Resing, J. (1997). Circular Markov chains. *The Mathematical Scientist* **22**, 27-31.

[2] Anderson, D.F., and Kurtz, T.G. (2015). *Stochastic Analysis of Biochemical Systems*. Springer International Publishing, Switzerland.

[3] Asmussen, S., and Koole, G. (1993). Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability* **30**, 365-372.

[4] Baccelli, F., Blaszczyszyn, B., and Karray, M. (2020). *Random Measures, Point Processes, and Stochastic Geometry*. INRIA, 2020, hal-02460214.

[5] Beentjes, C.H.L., Perez-Carrasco, R., and Grima, R. (2020). Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. *Physical Review E* **101**, 032403.

[6] Berg, O.G. (1978). A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of Theoretical Biology* **71**, 587-603.

[7] Bokes, P., King, J.R., Wood, A.T.A., and Loose, M. (2012). Exact and approximate distribution of protein and mRNA levels in the low-copy regime of gene expression. *Journal of Mathematical Biology* **64**, 829-854.

[8] Brémaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York.

[9] Bressloff, P.C. (2014). *Stochastic Processes in Cell Biology*. Springer International Publishing, Switzerland.

[10] Cao, Z., Filatova, T., Oyarzún, D.A., and Grima, R. (2020). A stochastic model of gene expression with polymerase recruitment and pause release. *Biophysical Journal* **119**, 1002-1014.

[11] Cao, Z., and Grima, R. (2020). Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proceedings of the National Academy of Sciences* **117**, 4682-4692.

[12] Choubey, S. (2018). Nascent RNA kinetics: transient and steady state behavior of models of transcription. *Physical Review E* **97** (2): 022402.

[13] Choubey, S., Kondev, J., and Sanchez, A. (2015). Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. *PLoS Computational Biology* **11**(11): e1004345.

[14] Dessalles, R., Fromion, V., and Robert, P. (2020). Models of protein production along the cell cycle: an investigation of possible sources of noise. *PLoS One* **15**(1), e0226016.

[15] Filatova, T., Popovic, N., and Grima, R. (2021). Statistics of Nascent and Mature RNA Fluctuations in a Stochastic Model of Transcriptional Initiation, Elongation, Pausing, and Termination. *Bulletin of Mathematical Biology* **83**, Issue 1, 1-62.

[16] Fromion, V., Leoncini, E., and Robert, P. (2013). Stochastic gene expression in cells: a point process approach. *SIAM Journal on Applied Mathematics* **73**, 195-211.

[17] Graham, R., Knuth, D., and Patashnik, O. (1994). *Concrete Mathematics*, Second Edition. Addison-Wesley, Upper Saddle River, NJ, USA.

[18] De Gunst, M., Mandjes, M. and Sollie, B. (2022). Statistical inference for a quasi-birth-death model of RNA transcription. *BMC Bioinformatics* **23**, 105.

[19] Ham, L., Schnoerr, D., Brackston, R.D., and Stumpf, M.P.H. (2020). Exactly solvable models of stochastic gene expression. *Journal of Chemical Physics* **152**, 144106.

[20] Ham, L., Jackson, M., and Stumpf, M.P.H. (2021). Pathway dynamics can delineate the sources of transcriptional noise in gene expression. *eLife* **10**, e69324.

[21] He, Q. (2014). *Fundamentals of Matrix-Analytic Methods*. Springer Science+Business Media, New York.

[22] Herbach, U. (2019). Stochastic gene expression with a multistate promoter: breaking down exact distributions. *SIAM Journal on Applied Mathematics* **79**, 1007-1029.

[23] Hornos, J.E.M., Schultz, D., Innocentini, G.C.P., Wang, J., Walczak, A.M., Onuchic, J.N., and Wolynes, P.G. (2005). Self-regulating gene: an exact solution. *Physical Review E* **72**, 051907.

[24] Horowitz, J.M., and Kulkarni, R.V. (2017). Stochastic gene expression conditioned on large deviations. *Physical Biology* **14**, 03LT01.

[25] Iyer-Biswas, S., Hayot, F., and Jayaprakash, C. (2009). Stochasticity of gene products from transcriptional pulsing. *Physical Review E* **79**, 031911.

[26] Jansen, M., and Pfaffelhuber, P. (2015). Stochastic gene expression with delay. *Journal of Theoretical Biology* **364**, 355-363.

[27] Jędrak, J., Kwiatkowski, M., and Ochab-Marcinek, A. (2019). Exactly solvable model of gene expression in a proliferating bacterial cell population with stochastic protein bursts and protein partitioning. *Physical Review E* **99**, 042416.

[28] Jia, C., Zhang, M.Q., and Qian, H. (2017). Emergent Lévy behavior in single-cell stochastic gene expression. *Physical Review E* **96**, 040402(R).

[29] Jia, C., Wang, L.Y., Yin, G.G., and Zhang, M.Q. (2019). Single-cell stochastic gene expression kinetics with coupled positive-plus-negative feedback. *Physical Review E* **100**, 052406.

[30] Jia, C. (2020). Kinetic foundation of the zero-inflated negative binomial model for single-cell RNA sequencing data. *SIAM Journal on Applied Mathematics* **80**, 1336-1355.

[31] Jia, C., and Grima, R. (2020). Small protein number effects in stochastic models of autoregulated bursty gene expression. *Journal of Chemical Physics* **152**, 084115.

[32] Jia, C., and Grima, R. (2021). Frequency domain analysis of fluctuations of mRNA and protein copy numbers within a cell lineage: theory and experimental validation. *Physical Review X* **11**, 021032.

[33] Jia, C., and Li, Y. (2022). Analytical time-dependent distributions for gene expression models with complex promoter switching mechanisms. Preprint accessible on bioRxiv at https://doi.org/10.1101/2022.01.05.475050

[34] Kallenberg, O. (1983). *Random Measures*. Akademie-Verlag, Berlin.

[35] Kelly, F.P. (1979). *Reversibility and Stochastic Networks*. Wiley.

[36] Ko, M.S.H. (1991). A stochastic model for gene induction. *Journal of Theoretical Biology* **153**, 181-194.

[37] Kumar, N., Platini, T. and Kulkarni, R.V. (2014). Exact distributions for stochastic gene expression models with bursting and feedback. *Physical Review Letters* **113**, 268105.

[38] Latouche, G., and Ramaswami, V. (1999). *Matrix-Analytic Methods in Stochastic Modeling*. ASA-SIAM publications, Philadelphia, PA, USA.

[39] Masuyama, M., and Takine, T. (2002). Analysis of an infinite-server queue with batch Markovian arrival streams. *Queueing Systems* **42**, 269-296.

[40] Paulsson, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews* **2**, 157-175.

[41] Peccoud, J. and Ycart, B. (1995). Markovian modeling of gene-product synthesis. *Theoretical Population Biology* **48**, 222-234.

[42] Perez-Carrasco, R., Beentjes, C., and Grima, R. (2020). Effects of cell cycle variability on linear and population measurements of messenger RNA abundance. *Journal of the Royal Society Interface* **17**, 20200360.

[43] Ramaswami, V., and Neuts, M. (1980). Some explicit formulas and computational methods for infinite-server queues with phase-type arrivals. *Journal of Applied Probability* **17**, 498-514.

[44] Rigney, D.R. (1979). Stochastic model of constructive protein levels in growing and dividing bacterial cells. *Journal of Theoretical Biology* **76**, 453-480.

[45] Rigney, D.R. (1979). Note on the kinetics and stochastics of induced protein synthesis as influenced by various models of messenger RNA degradation. *Journal of Theoretical Biology* **79**, 247-257.

[46] Rigney, D.R., and Schieve, W.C. (1977). Stochastic model of linear, continuous protein synthesis in bacterial populations. *Journal of Theoretical Biology* **69**, 761-766.

[47] Robert, P. (2019). Mathematical models of gene expression. *Probability Surveys* **16**, 277-332.

[48] Shahrezaei, V., and Swain, P.S. (2008). Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences* **105**, 17256-17261.

[49] Szavits-Nossan, J., and Grima, R. (2022). Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle. *Physical Review E* **105**, 014410.

[50] Zhou, T., and Zhang, J. (2012). Analytical results for a multistate gene model. *SIAM Journal on Applied Mathematics* **72**, 789-818.