Analytical solutions of the chemical master equation with bursty production and isomerization reactions

Gennady Gorin[1] and Lior Pachter[2]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125
[2]Division of Biology and Biological Engineering & Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125
[*]Address correspondence to Lior Pachter (lpachter@caltech.edu)

June 30, 2021

Contents

Abstract

Splicing cascades that alter gene products post-transcriptionally also affect expression dynamics. We study a class of processes and associated distributions that emerge from a bursty promoter model coupled to a path graph of downstream mRNA splicing, and more generally examine the behavior of finite-activity jump drivers coupled to a directed acyclic graph of splicing with one or more roots. These solutions provide full time-dependent joint distributions for an arbitrary number of species, offering qualitative and quantitative insights about how splicing can regulate expression dynamics. Finally, we derive a set of quantitative constraints on the minimum complexity necessary to reproduce gene co-expression patterns using synchronized burst models. We validate these findings by analyzing long-read sequencing data, where we find evidence of expression patterns largely consistent with these constraints.

1 Introduction

Recent advances in the analysis of single-cell RNA sequencing [1] and fluorescence microscopy [2] enable the quantification of pre-mRNA molecules alongside mature mRNA. These techniques provide an opportunity to infer the topologies and biophysical parameters governing the processes of mRNA transcription, splicing, export, and degradation in living cells. In particular, they provide a novel approach to inferring and studying the dynamics of splicing cascades [3].

However, drawing mechanistic conclusions from transcriptomics requires overcoming numerous statistical and computational challenges. Living cells contain mRNA in low copy numbers, and transient nascent species are even less abundant, leading to potential pitfalls if the discrete nature of the data is not appropriately modeled [4]. One approach to modeling dynamics from count data is to utilize detailed Markov models based on the chemical master equation (CME) [5–7]. Such modeling can, in principle, yield analytical solutions for the dynamics of genes affected by arbitrary splicing networks, thus circumventing tractability problems arising with matrix- and simulation-based methods that can be impractical for large numbers of species [8]. Several classes of analytical solutions to the CME are available [9], but their derivation tends to be *ad hoc*, with limited generalization to more complex systems.

Starting with the approach of Singh and Bokes to the problem of bursty transcription coupled to nuclear export and degradation of mRNA [10], we develop a class of solutions for gene dynamics affected by arbitrary splicing networks under the physiologically relevant assumption of bursty production of mRNA [11]. Fundamentally, the generating function of a Markov chain describing the

evolution of molecules in a splicing cascade can be automatically computed, numerically integrated in time, and then inverted by Fourier transformation at an overall computational complexity of $O(\mathcal{N} \log \mathcal{N})$ in state space size. We begin with the example of splicing described by a path graph, where the order of intron removal is deterministic, extend the procedure to a much broader class of splicing graphs and driving burst processes, and subsequently demonstrate the existence of the solutions and their isomorphism to a class of moving average processes.

2 Methods

2.1 Path graph splicing

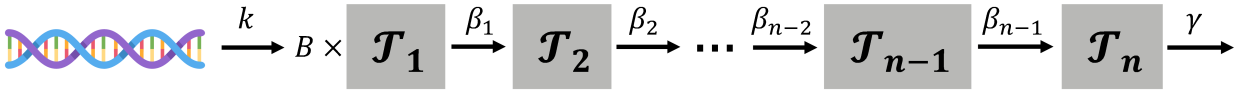

Figure 1: Graph representation of the generic path graph model. The source transcript $\mathcal{T}_1$ is synthesized at the gene locus in random geometrically distributed bursts (according to a distribution B with burst frequency k). Each molecule proceeds to isomerize in a chain of splicing reactions governed by successive rates $\beta_1, \beta_2, \ldots, \beta_{n-1}$, until reaching the form $\mathcal{T}_n$, which is ultimately degraded at rate γ.

Consider the system consisting of a bursting gene coupled to a n-step birth-death process, characterized by the path graph in Figure 1, where $B \sim Geom(b)$, and all reactions occur after exponentially-distributed waiting times. The bursts occur with rate k, the conversion of adjacent transcripts $\mathcal{T}_i$ to $\mathcal{T}_{i+1}$ occurs with rate β_i, and the degradation of $\mathcal{T}_n$ occurs with rate $\gamma := \beta_n$. We assume the rates of conversion and degradation are all distinct. The amount of species $\mathcal{T}_i$ can be described by the non-negative discrete random variable m_i. We assume no molecules are present at $t = 0$.

2.1.1 Discrete formulation and algorithm

We would like to compute the probability mass function (PMF) of the count distribution, denoted by $P(m_1, ..., m_n, t)$; this corresponds to the probability of observing m_1 molecules of $\mathcal{T}_1$, m_2 molecules of $\mathcal{T}_2$, etc., at time t. Following a previous derivation [10], this problem can be reframed in terms of a partial differential equation involving the probability generating function (PGF) $G(x_1, ..., x_n, t)$:

$$G := \sum_{m_1,\ldots,m_n} x_1^{m_1} \ldots x_n^{m_n} P(m_1, ..., m_n, t)$$

$$\frac{\partial G}{\partial t} = k(F(x_1) - 1)G + \sum_{i=1}^{n-1} \beta_i (x_{i+1} - x_i) \frac{\partial G}{\partial x_i} + \gamma(1 - x_n) \frac{\partial G}{\partial x_n},$$

where F is the PGF of the burst distribution. Applying the transformations $u_i := x_i - 1$ and $\phi := \ln G$ yields the equation:

$$\frac{\partial \phi}{\partial t} = k(M(u_1) - 1) + \sum_{i=1}^{n-1} \beta_i (u_{i+1} - u_i) \frac{\partial \phi}{\partial u_i} - \gamma u_n \frac{\partial \phi}{\partial u_n},$$

where $M(u) := F(1+u)$. This equation can be solved using the method of characteristics, with formal solution $\phi = \int_0^t [M(U_1(s)) - 1]ds$. The characteristics U_i, $i < n$ satisfy $\frac{dU_i}{ds} = \beta_i(U_{i+1} - U_i)$, with $U_n(u_1, ..., u_n; s) := U_n(s) = u_n e^{-\gamma s}$.
The functional form of $\frac{dU_i}{ds}$ implies that $U_1(s)$ is the weighted sum of exponentials $\sum_{i=1}^n A_i e^{-\beta_i s}$; the explicit analytical solution is provided in Section 5. The weights A_i can be computed through a simple iterative procedure, which proceeds from the terminal species and successively incorporates dependence on upstream rates:

$A_i \leftarrow 0 \; \forall \; i < n$
$A_n \leftarrow u_n$
$i \leftarrow n - 1$
while $i > 0$ **do**
  for $j > i$ **do**
    $A_j \leftarrow A_j \times \frac{\beta_i}{\beta_i - \beta_j}$
  end for
  $A_i \leftarrow u_i - \sum_{j>i} A_j$
  $i \leftarrow i - 1$
end while

Finally, given the physiologically plausible geometric burst distribution [12], $\phi(t;\cdot) = k\int_0^t \frac{bU_1(s)}{1-bU_1(s)}ds$. The stationary PGF is e^ϕ evaluated at $t = \infty$; this quantity can be converted to a probability mass function via Fourier inversion. This is an entirely general procedure that can be used for *any* path graph of downstream processing, although regions of stiffness are to be expected, particularly near the degenerate cases of matching rates. In the most general case, any β_i may be equal to any other, leading to a combinatorial explosion of auxiliary degenerate solutions. As the overarching motivation is the statistical inference of biophysical parameters, and the degenerate regions are all of measure zero, there appears to be little physical justification for considering them in further detail. However, we remark that these solution classes are likewise analytically tractable, with mixed polynomial-exponential functional forms entirely analogous to previous work [10].

2.2 Directed acyclic graph splicing

The simplicity of the ODEs governing the evolution of the CME lends itself to extensions to the broadest class of graphs representing the stochastic and incremental removal of discrete introns, the directed acyclic graphs.

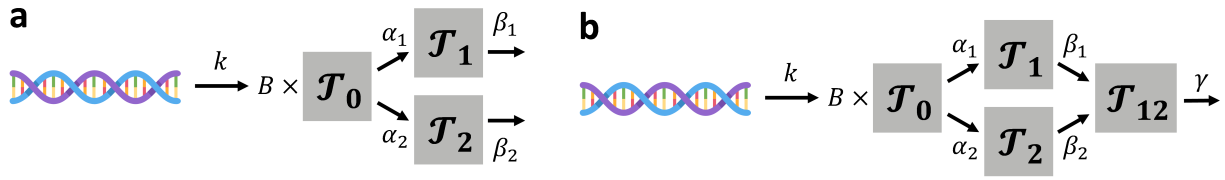

Figure 2: Graph representations of simple directed acyclic graph models. (a) Tree splicing graph with two terminal isoforms. (b) Convergent splicing graph with a single terminal isoform and two intermediate transcripts.

2.2.1 Example: alternative splicing

Suppose the downstream dynamics are given by a directed rooted tree. The solution procedure is analogous to that used for the path graph. First, starting at the leaves, the path subgraph solutions are produced by the procedure above, yielding a sum of exponentials. Then, at a node of out-degree > 1 (i.e., molecular species with several potential products), the associated ODE has a functional form identical to that of a path graph. Therefore, the solutions are analogous.
As an illustration, consider the simplest tree graph, shown in Figure 2a, where the splicing reactions occur at rates α_1 and α_2 and degradation reactions occur at rates β_1 and β_2. Physically, this graph can be interpreted as a single source mRNA being directly and stochastically converted to one of two terminal isoforms by removal of intron 1 or intron 2. Clearly, $U_i = u_i e^{-\beta_i s}$ for $i \in \{1, 2\}$.
The ODE governing the source species is $\frac{dU_0}{ds} = \alpha_1(U_1 - U_0) + \alpha_2(U_2 - U_0) = \alpha_1 U_1 + \alpha_2 U_2 - (\alpha_1 + \alpha_2)U_0$. The solution is $U_0 = \frac{\alpha_1}{\alpha_1+\alpha_2-\beta_1} u_1 e^{-\beta_1 s} + \frac{\alpha_2}{\alpha_1+\alpha_2-\beta_2} u_2 e^{-\beta_2 s} + K e^{-(\beta_1+\beta_2)s}$, with $K = u_0 - \frac{\alpha_1}{\alpha_1+\alpha_2-\beta_1} u_1 - \frac{\alpha_2}{\alpha_1+\alpha_2-\beta_2}$. Finally, the expression for U_0 can be directly plugged into the desired burst distribution generating function.

2.2.2 Example: two-intron splicing with non-deterministic order

Consider the same tree graph as in the example above, and suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ are converted to product $\mathcal{T}_{12}$ at rates β_1 and β_2, as shown in Figure 2b. Afterward, $\mathcal{T}_{12}$ is degraded at rate γ. Physically, this graph can be interpreted as a single source mRNA being converted to one of two intermediate isoforms by the removal of one of two introns, then to a single terminal isoform by the removal of the other intron. Clearly, $U_{12} = u_{12} e^{-\gamma s}$. Setting $f_i := \frac{\beta_i}{\beta_i - \gamma}$, we find $U_i = (u_i - f_i u_{12}) e^{-\beta_i s} + f_i u_{12} e^{-\gamma s}$. Finally, the dynamics of the source molecule $\mathcal{T}_0$ are governed by the following ODE:

$$\frac{dU_0}{ds} = \alpha_1(u_1 - f_1 u_{12})e^{-\beta_1 s} + \alpha_2(u_2 - f_2 u_{12})e^{-\beta_2 s} + (\alpha_1 f_1 + \alpha_2 f_2)u_{12} e^{-\gamma s} - (\alpha_1 + \alpha_2)U_0$$

Yet again, the functional form affords a straightforward analytical solution:

$$U_0 = K e^{-cs} + \frac{C_1}{c - \beta_1} e^{-\beta_1 s} + \frac{C_2}{c - \beta_2} e^{-\beta_2 s} + \frac{C_3}{c - \gamma} e^{-\gamma s},$$

where $c := \alpha_1 + \alpha_2$, $C_1 := \alpha_1(u_1 - f_1 u_{12})$, $C_2 := \alpha_2(u_2 - f_2 u_{12})$, and $C_3 := \alpha_1 f_1 + \alpha_2 f_2$. From the initial condition $U_0(s = 0) = u_0$, we yield $K = u_0 - \frac{C_1}{c-\beta_1} - \frac{C_2}{c-\beta_2} - \frac{C_3}{c-\gamma}$. The details of the ODE solution are described in Section 5 and the computation procedure is demonstrated in Figure 3.

2.2.3 Algorithm

This procedure is facile to extend to an arbitrary directed acyclic graph with a unique root. The reaction system is fully characterized by two arrays, the stoichiometric matrix matrix S and the rate vector r used in the stochastic simulation algorithm [13]. Thus, for example, the path graph

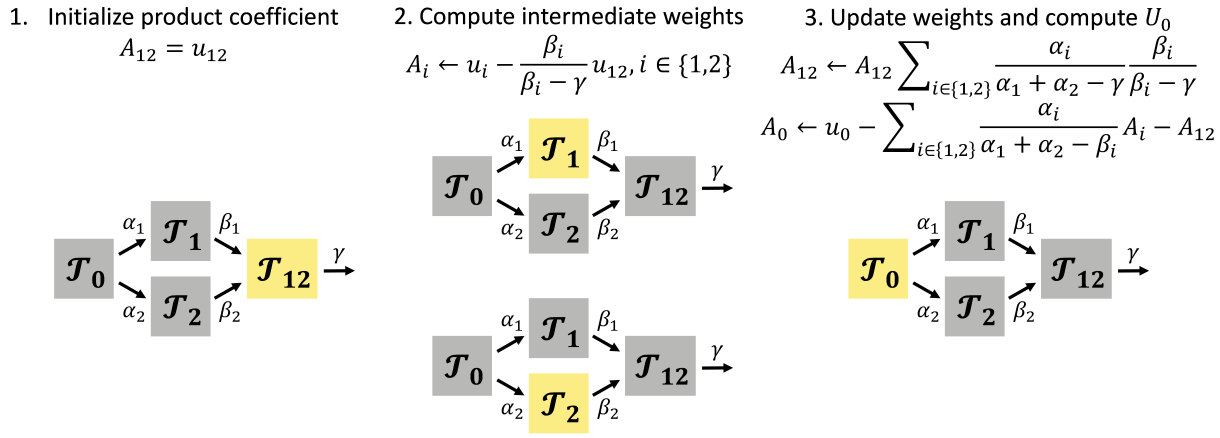

Figure 3: Illustration of the solution algorithm. The differential equation structure requires the backward propagation of downstream species' solutions, weighted by ratios of rates.

can be represented by the full rate vector $[k, \beta_1, ..., \beta_{n-1}, \gamma]$ and the following stoichiometric matrix:

$$\begin{bmatrix} B & 0 & 0 & ... & 0 & 0 \\ -1 & 1 & 0 & ... & 0 & 0 \\ 0 & -1 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -1 & 1 \\ 0 & 0 & 0 & ... & 0 & -1 \end{bmatrix}$$

A well-defined system has n species and $q + 1$ reactions. We assume unbounded accumulation does not occur and the graph is connected, so each species has influx and efflux reaction pathways, implying $q \geq n$. Further, we suppose that each species is associated with at most one degradation reaction; if several channels are available, they can be added by the superposition property of a Poisson process. Finally, we assume that all isomerization and degradation rates are distinct. Since the underlying graph is acyclic, there exists at least one terminal species with a single efflux reaction. Thus, consider a single production reaction with rate k coupled to a set of isomerization reactions with rates c_{ji} and degradation reactions with rates c_{j0}.

The downstream dynamics are determined by U_1, an exponential sum with n terms. In this generic case, *lumped* rate exponents r_i must be computed as the net efflux rate from each molecule. A simple implementation of the routine is outlined below:

$r_i \leftarrow \sum_j c_{ij} \ \forall \ i$ ▷ Compute all exponents
$A_{ij} \leftarrow 0 \ \forall \ i, j \in [1, n]$ ▷ Initialize coefficient array
$R_D \leftarrow \{k \mid \sum_i S_{ki} = -1\}$ ▷ Identify reaction channels that correspond to degradation
$D \leftarrow \{i \mid S_{ki} = -1 \text{ for } k \in R_D\}$ ▷ Identify degraded species
$T \leftarrow \{j \in D \mid \nexists \ (S_{kj} = -1 \cap S_{ki} = 1)\}$ ▷ Identify terminal species
$R_T \leftarrow \{k \mid S_{ki} = -1 \text{ for } i \in T\}$ ▷ Identify terminal reactions
$A_{ii} \leftarrow u_i \ \forall \ i \in T$ ▷ Assign terminal degradation coefficients
$C \leftarrow \{\varnothing\}$ ▷ Initialize set of computed species

while $|C| < n$ **do**
  for $j \notin C$ **do** ▷ Iterate over all species that have not yet been computed
    $R_I = \{k \mid (S_{kj} = -1 \cap S_{ki} = 1)\}$ ▷ Identify isomerization reactions originating at j
    $P = \{i \mid (S_{ki} = 1 \text{ for } k \in R_I)\}$ ▷ Identify corresponding isomerization products
    if $P \subset C$ **then** ▷ If all products have been computed
      for $i \in P$ **do** ▷ For each isomerization product
        $A_{jk} \leftarrow A_{jk} + A_{ik} \times \frac{c_{ji}}{r_j - r_i}$ ▷ Propagate results from downstream solutions
      end for
      $A_{jj} \leftarrow u_j - \sum_k A_{jk}$ ▷ Apply initial condition, compute coefficient for $e^{-r_j s}$
      $C \leftarrow C \cup j$ ▷ Adjust set of computed species
    end if
  end for
end while

The terminal exponential sum is given by $U_1 = \sum_{i=1}^{n} A_{1i} e^{-r_i s}$. As before, the full time-dependent solution for any burst distribution with a well-defined moment-generating function is computable by quadrature. The case of the geometric burst distribution yet again corresponds to $\phi(t; u_1, ..., u_n) = k \int_0^t \frac{bU_1(s)}{1-bU_1(s)} ds$.

2.3 Continuous formulation

We have defined a series of discrete joint distributions induced by a graph governing splicing and degradation. However, we can equivalently recast the problem in terms of the stochastic differential equations (SDEs) governing the distributions' Poisson intensities Λ_i. Specifically, the following identity can be used to relate a set of stochastic processes $\Lambda_1, ..., \Lambda_n$ with joint distribution $F_{\Lambda_1,...,\Lambda_n}$ to the solution of the CME [9, 14–16]:

$$P(m_1, ..., m_n, t) = \int \prod_{i=1}^{n} \frac{e^{-\Lambda_i} \Lambda_i^{m_i}}{m_i!} dF_{\Lambda_1,...,\Lambda_n}$$

As an illustration, we can consider the intensity of an n-step isomerization process driven by a finite-activity compound Poisson Lévy subordinator L_t:

$$\begin{aligned} d\Lambda_1 &= -\beta_1 \Lambda_1 dt + dL_t \\ d\Lambda_2 &= -\beta_2 \Lambda_2 dt + \beta_1 \Lambda_1 dt \\ &... \\ d\Lambda_n &= -\gamma \Lambda_n dt + \beta_{n-1} \Lambda_n dt, \end{aligned}$$

This formulation has an intimate connection with the theory of moving average processes, which can be immediately seen by applying variation of parameters to the Poisson representation:

$$\Lambda_i^p = Ce^{-\beta_i t}$$
$$\Lambda_i = C_t e^{-\beta_i t}$$
$$d\Lambda_i = -\beta_i C_t e^{-\beta_i t} dt + e^{-\beta_i t} dC_t$$
$$= -\beta_i C_t e^{-\beta_i t} dt + \beta_i \Lambda_{i-1} dt$$
$$C_t = \int_0^t \beta_i \Lambda_{i-1}(s) e^{\beta_i s} ds$$
$$\Lambda_i(t) = e^{-\beta_i t} C_t = \int_0^t \beta_i \Lambda_{i-1}(s) e^{-\beta_i (t-s)} ds$$

By convention, $\beta_n := \gamma$ and $\Lambda_0 ds := dL_s$, the underlying driving subordinator.
Thus, each Λ_i is the exponentially-weighted moving average of Λ_{i-1}; Λ_1 is the moving average of the Poisson shot noise introduced by the compound Poisson subordinator. Although ample literature exists on the topic of moving average processes [17], it generally considers the problem of inference and prediction from discrete-time observations. Furthermore, discussions in the context of the chemical systems only tend to consider the first-order moving average of shot noise [9].
This formulation implies that any n-step isomerization process is identical to an $(n-i)$-step isomerization process driven by an order i iterated moving average process. This identity affords a route for the analytical computation of hybrid continuous-discrete system solutions merely by marginalizing the first i species.
The Poisson representation enables the simultaneous discussion of the properties of $F_{m_1,\ldots,m_n}$, the stationary distribution of the continuous-time Markov chain, and $F_{\Lambda_1,\ldots,\Lambda_n}$, the stationary distribution of the underlying series of stochastic differential equations. In fact, from standard properties of Poisson mixtures [18], the PGF $G(x_1,\ldots,x_n)$ of the former evaluated at $u_i = x_i - 1$, precisely yields the moment-generating function (MGF) of the latter.
This connection also allows the evaluation of the solution for a broad class of compound Poisson driving processes. Specifically, $M(u) = \frac{1}{1-bu}$ is the MGF associated with exponentially distributed jumps. In the discrete domain, this translates to the geometric burst distribution. However, the downstream dynamics, represented by $U_1(s)$, are independent of the burst distribution. Therefore, the procedure affords computable solutions for any finite-activity pure-jump Lévy driving process. Extensions to generic directed acyclic graphs are entirely analogous. These solutions, identical to the CME case, can be used to exactly solve arbitrary DAGs representing *continuous*-valued molecular concentrations. This representation is conventional for high-abundance species [19].
This formulation occasionally enables the confirmation of CME results using standard properties of SDEs. For example, when L_t is a compound Poisson subordinator with exponential jumps, Λ_1 is the gamma Ornstein–Uhlenbeck process [20–23]. Inspired by a highly general result for constitutive transcription, which states that Poisson distributions always remain Poisson for a birth-death process [14], we may reasonably ask whether equivalent results are available for bursty processes. This intuition turns out incorrect, and straightforward to disconfirm using SDE results. The distribution of the gamma Ornstein–Uhlenbeck process is not gamma for any finite $t \in (0,\infty)$ – although it does approach a gamma law exponentially fast [21]; therefore, the corresponding Poisson mixture is *not* simply a time-varying negative binomial.

2.4 Distribution properties

Using the algorithms above, we can compute the generating functions corresponding to the stochastic processes of interest. However, it is not yet clear that these generating functions are everywhere well-defined. For example, certain physiologically plausible noise models do not have all moments [24]; their generating functions fail to converge in certain regimes. By analyzing the functional form of the downstream process and assuming geometric-distributed bursts, we demonstrate that this class of processes is guaranteed to yield convergent generating functions and finite moments.

2.4.1 Positivity of the exponential sum

First, we demonstrate that the downstream processes yield a strictly positive functional form of time dependence. Noting that the marginal of species i yields the functional form $U_1(u_i; s) = u_i \sum_{j=1}^{n} a_j e^{-r_j s} := u_i \psi_i(s)$, this condition translates to $\psi_i(s) > 0$ for all $s > 0$.
Consider $F(x) = x$, corresponding to constitutive production of the source species (i.e. a Poisson birth process), with no molecules present at $t = 0$. Focusing on the marginal of species i, this assumption yields $\phi(u_i; t) = k \int_0^t U_1(u_i; s)ds = k \int_0^t u_n \psi_n(s)ds$. Evaluating e^{ϕ} at $x_i = 0$, i.e. $u_i = -1$, marginalizes over all $j \neq i$ and yields the probability of observing zero counts of species i: $G(u_i; t) = P(m_i = 0, t) = P_0(t) = \exp(-k \int_0^t \psi_i(s)ds)$. The corresponding time derivative is $\frac{dP_0}{dt} = -k\psi_n(t) \exp(-k \int_0^t \psi_i(s)ds)$. Simultaneously, we know that $P_0(t) = e^{-\lambda_i(t)}$, where $\lambda_i(t)$ is the solution of the reaction rate equation for species i [14]. Clearly, $\frac{dP_0}{dt} = -P_0(t)\frac{d\lambda_i(t)}{dt}$. The reaction rate $\frac{d\lambda_i(t)}{dt} > 0$ at $t = 0$ under the given initial conditions. Furthermore, $\frac{d\lambda_i(t)}{dt}$ is strictly positive. This follows from the reaction rate equations. By the continuous formulation, λ_i is a weighted moving average of some set of processes $\{\lambda_k\}$. λ_1 is a strictly increasing function governed by $\frac{k}{r_1}(1 - e^{-r_1 s})$. The property of being strictly increasing is retained under moving average and rescaling. Therefore, each successive moving average must be strictly increasing.
Finally, $P_0 \in (0, 1) > 0$, because the solution for m_i is given by a Poisson distribution, which has support on all of $\mathbb{N}_0$. Therefore, $\frac{dP_0}{dt}$ is strictly negative. As the exponential term and k are positive, this implies $\psi_i(s)$ is strictly positive for all $s > 0$.

2.4.2 Existence of generating functions and moments

Next, we show that $G(u_i; t)$, the generating function of the ith marginal, is finite for the geometric burst system. This follows from the construction of the original PGF: the marginal PGF is guaranteed to be finite if $1 - bu_i\psi_i$ is never zero. But for the relevant domain $\Re(u_n) \leq 0$, on the shifted complex unit circle, $\Re(1 - bu_i\psi_i) \geq 1$, except at the degenerate initial case. The existence of the marginal moments of m_i is implied by the existence of the generating function. The existence of all cross moments follows from the Cauchy-Schwartz inequality. Per standard properties, this existence property holds for both m_i and Λ_i.
The tails of the stationary discrete marginals decay no slower than the geometric distribution. This follows immediately from the lower bound on $\Re(1 - bu_i\psi_i)$, which in turn gives an upper bound on x_i [10]. Equivalently, this follows from the existence of all moments [24]. An analytical radius of convergence has been given previously for $n = 2$ [10], but numerical optimization is necessary to establish rates of tail decay for $n > 2$.

2.4.3 Statistical properties

All marginals are infinitely divisible. This follows from the functional form of the PGF: the random variable corresponding to any marginal distribution can be written as as a sum of q random variables with burst frequency k/q.

Only the first marginal is self-decomposable. This follows from the condition that a random variable has a self-decomposable (sd) law if and only if it offers a representation of the form $Y = \int_0^\infty e^{-t} dX_t$, with Lévy X_t [25]. However, only Λ_0 is Lévy. All downstream intensity processes have nontrivial, almost-everywhere C^∞ trajectories, which implies they cannot be represented by a Lévy triplet: the only permitted continuous Lévy processes are linear combinations of the (non-differentiable) Brownian motion W_t and the trivial process t. Therefore, Λ_i is sd for $i = 1$ and non-sd for all $i > 1$.

All stationary marginals are unimodal. Multimodality in the distribution of the moving average over time is contingent on time-inhomogeneity in the trajectory process. However, the underlying driving process is defined to be time-homogeneous, with uniformly distributed jump times; therefore, each downstream process has a unimodal distribution over time. By ergodicity, this distribution is equivalent to the ensemble distribution. Therefore, all marginals of downstream species are unimodal.

2.4.4 Moments

The moments of the marginals can be computed directly from the derivatives of the marginal MGF $e^{\phi(u_i)}$ at $u_i = 0$:

$$\frac{de^{\phi}}{du_i} = e^{\phi}\frac{d\phi}{du_i} = e^{\phi}\frac{d}{du_i}k\int_0^\infty \frac{bu_i\psi_i(s)}{1 - bu_i\psi_i(s)}ds$$
$$= kb\int_0^\infty \psi_i(s)ds = kb\int_0^\infty \sum_{i=1}^n a_i e^{-r_i s}ds = kb\sum_{i=1}^n \frac{a_i}{r_i}.$$

For the path graph system, the CME immediately yields $\mu_i = \frac{kb}{r_i}$. Considering the previously discussed constitutive solution $\phi^0(u_i) = ku_i\int_0^\infty \psi_i(s)ds$, we yield the identity $\int_0^\infty \psi_i(s)ds = \sum_{i=1}^n \frac{a_i}{r_i}$, which is equal to $\frac{1}{r_i}$ for the path graph system. Per the standard properties of mixed Poisson distributions [26], the value of μ_i is identical for the underlying continuous process and the derived discrete process.
The stationary second moments can be found analogously:

$$\frac{d^2e^{\phi}}{du_i^2} = e^{\phi}\frac{d^2\phi}{du_i^2} + e^{\phi}\left(\frac{d\phi}{du_i}\right)^2$$
$$= 2kb^2\int_0^{\infty}\psi_i^2(s)ds + \left(\frac{kb}{r_i}\right)^2$$
$$\psi_n^2(s) = \sum_{j,k=1}^{n} a_j a_k e^{-(r_k+r_k)s} = \sum_{j=1}^{n} a_j^2 e^{-2r_j s} + \sum_{j,k=1,\ j\neq k}^{n} a_j a_k e^{-(r_j+r_k)s}$$
$$\frac{d^2e^{\phi}}{du_i^2} = 2kb^2\sum_{j,k=1}^{n}\frac{a_j a_k}{r_j+r_j} + \mu_i^2 = \mathbb{E}[\Lambda_i^2]$$
$$\mathbb{V}[\Lambda_i] = 2kb^2\sum_{j,k=1}^{n}\frac{a_j a_k}{r_j+r_k},$$

which is straightforward to compute, but does not appear to have an easily amenable analytical form beyond the first few cases. Naturally, the standard properties of Poisson mixtures [26] allow conversion to the discrete domain, with $\mathbb{V}[m_i] = \mathbb{V}[\Lambda_i] + \mu_i$.

The covariances can be computed directly from the derivatives of the MGF $e^{\phi(u_l,u_i)}$, i.e., the marginalized MGF for $u_q = 0$ for all $q \neq l, i$. The particular functional form of $\phi(u_l, u_i)$ is given by $k\int_0^{\infty}\frac{bU_1(u_i,u_l;s)}{1-bU_1(u_i,u_l;s)}$. By construction, $U_1(u_l, u_i; s) = u_l\psi_l(s)+u_i\psi_i(s)$, where each ψ is the exponential sum corresponding to the marginal of the species in question. This yields:

$$\frac{d^2e^{\phi}}{du_l du_i} = \frac{d}{du_l}e^{\phi}k\int_0^{\infty}\frac{b\psi_i ds}{(1-bu_l\psi_l - bu_i\psi_i)^2}$$
$$= e^{\phi}k^2\int_0^{\infty}\frac{b\psi_l ds}{(1-bu_l\psi_l - bu_i\psi_i)^2}\int_0^{\infty}\frac{b\psi_i ds}{(1-bu_l\psi_l - bu_i\psi_i)^2}$$
$$+ 2e^{\phi}k\int_0^{\infty}\frac{b^2\psi_l\psi_i ds}{(1-bu_l\psi_l - bu_i\psi_i)^3}$$
$$\mathbb{E}[\Lambda_l\Lambda_i] = k^2b^2\int_0^{\infty}\psi_l ds\int_0^{\infty}\psi_i ds + 2kb^2\int_0^{\infty}\psi_l\psi_i ds$$
$$= \mu_l\mu_i + 2kb^2\int_0^{\infty}\psi_l\psi_i ds = \mathrm{Cov}(\Lambda_l,\Lambda_i) + \mu_l\mu_i,$$

which implies that

$$\mathrm{Cov}(\Lambda_l,\Lambda_i) = 2kb^2\int_0^{\infty}\psi_l\psi_i ds.$$

As above,

$$\psi_l(s)\psi_i(s) = \sum_{j,k=1}^{n} a_j c_k e^{-(r_j+r_k)s}$$

$$\int_0^\infty \psi_l(s)\psi_i(s)ds = \sum_{j,k=1}^{n} \frac{a_j c_k}{r_j + r_k}$$

$$\mathrm{Cov}(\Lambda_l, \Lambda_i) = 2kb^2 \sum_{j,k=1}^{n} \frac{a_j c_k}{r_j + r_k} = 2kb^2 \sum_{j,k=1}^{n} \frac{a_j c_k}{r_j + r_k},$$

where a_j are the weights associated with ψ_l and c_k are the weights associated with ψ_i. This form of the summation is very general; for example, in case of the path graph system with $l > i$, it can be equivalently represented as $2kb^2 \sum_{j=1}^{l} \sum_{k=1}^{i} \frac{a_j c_k}{r_j+r_k}$. From standard identities, the covariance of a mixed bivariate Poisson distribution with no intrinsic covariance forcing is identical to the covariance of the mixing distribution [26]. Therefore, this result holds for both the CME and the underlying SDE.

The Pearson correlation coefficient follows immediately from the result above:

$$\rho = \frac{\sum_{j,k=1}^{n} \frac{a_j c_k}{r_j+r_k}}{\sqrt{\sum_{j,k=1}^{n} \frac{a_j a_k}{r_j+r_k} \times \sum_{j,k=1}^{n} \frac{c_j c_k}{r_j+r_k}}}$$

Since mixing decreases variance but not covariance, the correlation coefficient of the discrete system will always be lower than that of its continuous or hybrid analog.

2.5 Simulation

To compare the analytical solutions with simulation, we generated a random directed acyclic graph, shown in Figure 4. The numbers of species (7) and isomerization reactions (11) were chosen arbitrarily. We enforced the existence of a single unique source node (a) and the weakly connected property to ensure only a single source mRNA would be present and all isoforms would be reachable from it, but did not impose any other conditions. The number of degraded species (3) was chosen arbitrarily; we assigned degradation reactions to the two sink species (c, e) and randomly chose a degraded intermediate (b) from a uniform distribution over the molecular species.

All reaction rates were drawn from a log-uniform distribution on $[10^{-0.5}, 10^{0.5}]$; we chose to sample them from a single order of magnitude to avoid the trivial degenerate cases that occur in cases of very slow or very fast export [10]. This process produced the parameter values $k = 0.44$, $\beta = [0.48, 2.12, 1.31, 2.21, 1.16, 2.41, 0.4, 1.19, 0.37, 1.19, 0.53]$, and $\gamma = [0.94, 2.38, 0.72]$, with the indices corresponding to those in Figure 4. Finally, we chose the geometric burst model with $b = 10$.

We applied the algorithm to compute the exponents and coefficients, and computed the stationary distributions of all species. The simulated distributions match the quantitative results for the marginals, as shown in Figure 5. Furthermore, the 49 entries of the covariance matrix are likewise effectively predicted by the procedure for moment calculation.

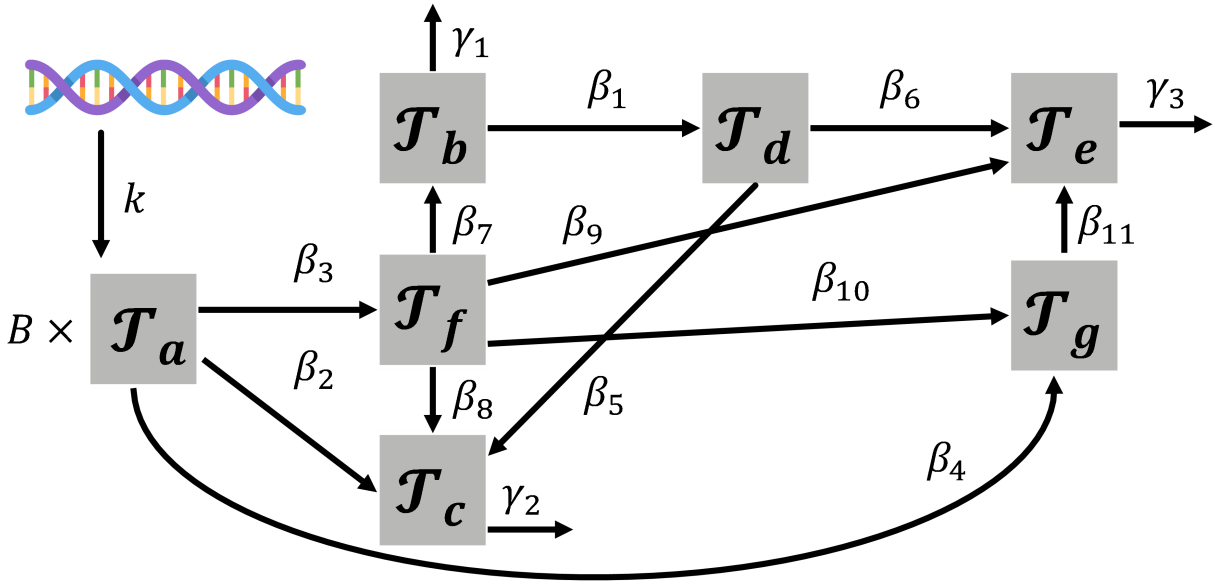

Figure 4: Graph representation of the randomly generated transcription, splicing, and degradation model. A single source isoform $\mathcal{T}_a$ is converted to a variety of downstream isoforms $\mathcal{T}_b, ..., \mathcal{T}_g$, which isomerize according to a directed acyclic graph.

2.6 Multi-gene systems

Although the CME solution is a general framework, the relevance to modern transcriptomic experimental data is tempered somewhat by the simplicity of the model. The model can describe the splicing cascade of a single gene, but does not naturally extend to multi-gene networks. Yet we know that genes often belong to co-expression modules that are identifiable by similarity metrics [3, 27]. Therefore, we are faced with the challenge of integrating multiple genes in a physically meaningful way.

Instead of building intractable "top-down" models that encode complex networks, we may build "bottom-up" models that extend analytical solutions. For example, we can consider sets of *synchronized* genes that experience bursting events at the same time. This model represents the bursty limit of multiple genes with transcription rates governed by a single telegraph process, up to scaling; a conceptually similar model has previously been used to describe correlations between multiple copies of one gene [28]. This model retains the appeal of physical interpretability – for example, gene modules may be regulated by the same molecule – but does not excessively complicate the mathematics, and offers an incremental step toward more detailed descriptions.

2.6.1 Two-gene bursty model, no splicing

We begin by considering the instructive model of two genes influenced by the same regulator, with no downstream splicing. The burst processes are synchronized, but the burst sizes are not, and may indeed come from different distributions.

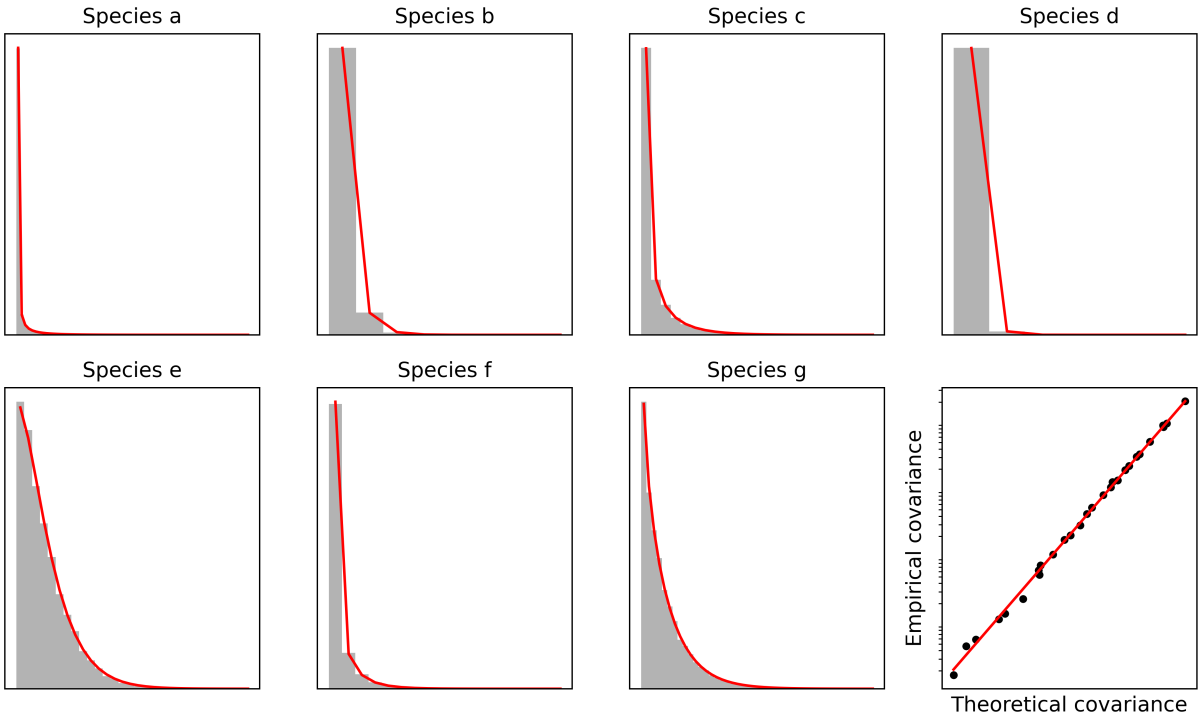

Figure 5: Simulation of the randomly generated graph model. Histograms: empirical results (10,000 simulations). Black dots: theoretical and empirical covariance. Red lines: theory.

$$\varnothing \xrightarrow{k} B_1 \times \mathcal{T}_1 + B_2 \times \mathcal{T}_2$$
$$\mathcal{T}_1 \xrightarrow{\beta_1} \varnothing$$
$$\mathcal{T}_2 \xrightarrow{\beta_2} \varnothing$$

The following CME holds:

$$\begin{aligned}\frac{dP(m_1,m_2,t)}{dt} = k &\left(\sum_{i=0}^{m_1}\sum_{j=0}^{m_2} p_i q_j P(m_1-i, m_2-j, t) - P(m_1,m_2,t)\right) \\ &- \beta\left((m_1+1)P(m_1+1,m_2,t) - m_1 P(m_1,m_2,t)\right) \\ &- \gamma\left((m_2+1)P(m_1,m_2+1,t) - m_2 P(m_1,m_2,t)\right),\end{aligned}$$

where p_i and q_j give the PMF weights of the burst size distributions that govern B_1 and B_2. We define the joint PGF

$$G(x_1,x_2,t) := \sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty} P(m_1,m_2,t) x_1^{m_1} x_2^{m_2},$$

and recognize that the degradation terms have the familiar functional forms $-\beta_1(x_2-1)\frac{\partial G}{\partial x_1}$ and $-\beta_2(x_2-1)\frac{\partial G}{\partial x_2}$. Therefore, considering the burst term:

$$\begin{aligned}
&\sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty} x_1^{m_1}x_2^{m_2}\sum_{i=0}^{m_1}\sum_{j=0}^{m_2} p_i q_j P(m_1-i, m_2-j, t)\\
&= \sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty}\sum_{i=0}^{m_1}\sum_{j=0}^{m_2} x_1^{m_1}x_2^{m_2} p_i q_j P(m_1-i, m_2-j, t)\\
&= \sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty}\sum_{i=0}^{n}\sum_{j=0}^{m}(x_1^i p_i)(x_2^j q_j) x_1^{m_1-i} y^{m_2-j} P(m_1-i, m_2-j, t)\\
&= \sum_{m_1=0}^{\infty}\sum_{i=0}^{n}(x_1^i p_i) x_1^{m_1-i}\sum_{m_2=0}^{\infty}\sum_{j=0}^{m_2}(x_2^j q_j) x_2^{m_2-j} P(m_1-i, m_2-j, t)\\
&= \sum_{m_1=0}^{\infty}\sum_{i=0}^{m_1}(x_1^i p_i) x_1^{m_1-i} F_2(x_2) G'(x_2; m_1-i)\\
&= F_1(x_1)F_2(x_2)G(x_1, x_2),
\end{aligned}$$

where G' is the conditional PGF assuming $m_1 - i$ molecules of $\mathcal{T}_1$ and F_i are the burst distribution PGFs; the final steps exploit the interpretation of the double sums as Cauchy products. This result implies:

$$\frac{\partial G}{\partial t} = k(F_1(x_1)F_2(x_2)-1)G - \beta_1(x_1-1)\frac{\partial G}{\partial x_1} - \beta_2(y_2-1)\frac{\partial G}{\partial y_2};$$

defining $G = e^{\phi}$, $u_i = x_i - 1$, and $M_i(u_i) = F_i(1+u_i)$:

$$\frac{\partial \phi}{\partial t} = k(M_1(u_1)M_2(u_2)-1) - \sum_{i=1}^{2}\beta_i u_i \frac{\partial \phi}{\partial u_i}.$$

The PDE can be solved using the method of characteristics, yielding $U_i = u_i e^{-\beta_1 s}$:

$$\begin{aligned}
\phi = k\int_0^t \left[M_1(U_1)M_2(U_2)-1\right] ds &= k\int_0^t \left[\frac{1}{1-b_1U_1}\frac{1}{1-b_2U_2} - 1\right] ds\\
&= k\int_0^t \left[\frac{1}{1-b_1u_1e^{-\beta_1 s}}\frac{1}{1-b_2u_2e^{-\beta_2 s}} - 1\right] ds,
\end{aligned}$$

assuming, as before, that $B_i \sim Geom(b_i)$.

2.6.2 n-gene bursty model with splicing

From the derivation above, four interesting properties stand out. Firstly, the functional form of U_i depends only on the details of the downstream processes; as shown, it can be easily computed for any DAG. Secondly, the derivation holds with no loss of generality for *any* number of genes, so long as the burst distributions are uncoupled, by repeated application of the Cauchy product formula.

Thirdly, even if they *are* coupled, the derivation still holds, but with intermediate conditional PGFs F_i, ultimately yielding a joint PGF F that factorizes in the independent case. Finally, even though we have considered the case of disjoint DAGs, the results still hold if the gene products can ultimately converge. This becomes self-evident by identifying $\mathcal{T}_1$ with $\mathcal{T}_2$ and setting $u_1 = u_2$, $b_1 = b_2$, and $\beta_1 = \beta_2$: the synchronized loci are merely multiple copies of the same gene, and produce bursts sampled from a *negative binomial* distribution, the sum of two identical geometric distributions. Therefore, in the most general case of arbitrary burst distributions and downstream splicing cascades, the factorial-cumulant generating function takes the following form:

$$\phi = k \int_0^t \left[M(U_1, ..., U_N) - 1 \right] ds,$$

where $U_1, ..., U_N$ are now the "U_1" functions, *not necessarily distinct*, corresponding to the species produced by each bursty locus, and M is the joint MGF of the burst distributions. The single-gene results are recovered by marginalization.

2.6.3 Examples of multi-gene systems

Uncorrelated gene loci By the Cauchy-Schwartz inequality, all moments and cross-moments exist. The correlations are straightforward to compute. For example, we can consider the stationary covariance of gene products from two co-expressed genes, with disjoint downstream splicing processes. We marginalize over all other transcripts and consider the functional forms $U_l = u_l \psi_l$ and $U_i = u_i \psi_i$, supposing that the respective average burst sizes are b_l and b_i.

$$\begin{aligned}
\frac{d^2 e^\phi}{du_l du_i} &= \frac{d}{du_l} e^\phi k \int_0^\infty \frac{1}{(1 - b_l u_l \psi_l)} \frac{b_i \psi_i ds}{(1 - b_i u_i \psi_i)^2} \\
&= e^\phi k^2 \int_0^\infty \frac{1}{1 - b_l u_l \psi_l} \frac{b_i \psi_i ds}{(1 - b_i u_i \psi_i)^2} \int_0^\infty \frac{b_l \psi_l ds}{(1 - b_l u_l \psi_l)^2} \frac{1}{1 - b_i u_i \psi_i} \\
&+ e^\phi k \int_0^\infty \frac{b_l \psi_l}{(1 - b_l u_l \psi_l)^2} \frac{b_i \psi_i ds}{(1 - b_i u_i \psi_i)^2} \\
\mathbb{E}[\Lambda_l \Lambda_i] &= k^2 b_i b_l \int_0^\infty \psi_l ds \int_0^\infty \psi_i ds + k b_i b_l \int_0^\infty \psi_l \psi_i ds \\
&= \mu_l \mu_i + k b_i b_l \int_0^\infty \psi_l \psi_i ds = \mathrm{Cov}(\Lambda_l, \Lambda_i) + \mu_l \mu_i,
\end{aligned}$$

which implies that

$$\mathrm{Cov}(\Lambda_l, \Lambda_i) = k b_i b_l \int_0^\infty \psi_l \psi_i ds = k b_i b_l \sum_{j,k=1}^{n} \frac{a_j c_k}{r_j + r_k},$$

where a_j are the weights associated with ψ_l and c_k are the weights associated with ψ_i, much as before. The covariance of the discrete process is identical. We reiterate that this expression *only* applies when the splicing graphs downstream of the gene loci are disjoint. The converse case requires slightly more unwieldy notation to account for, e.g., species l accessible from N source transcripts being associated with distinct coefficients $a_l^{(1)}, a_l^{(2)}, ..., a_l^{(N)}$, and does not appear to have a simple analytical expression totally agnostic of the number of loci and accessibility; nevertheless, it is easily

tractable by appropriately defining the function sets $\{U_l^{(j)}\}$ and $\{U_i^{(j)}\}$ and using the procedure above.

With this solution in hand, we can revisit the two-gene problem with no splicing:

$$\mathrm{Cov}(\Lambda_1, \Lambda_2) = \frac{k b_1 b_2}{\beta_1 + \beta_2}. \tag{1}$$

We are interested in standard summaries, such as the Pearson correlation coefficient:

$$\begin{aligned}
\rho = \frac{\mathrm{Cov}(\Lambda_1, \Lambda_2)}{\sigma_1 \sigma_2} &= \frac{k b_1 b_2}{\beta_1 + \beta_2} \times \sqrt{\frac{\beta_1 \beta_2}{k^2 b_1 (1+b_1) b_2 (1+b_2)}} \\
&= \frac{\sqrt{\beta_1 \beta_2}}{\beta_1 + \beta_2} \times \sqrt{\frac{1}{(1+1/b_1)(1+1/b_2)}} \\
&= \frac{\sqrt{\beta_1/\beta_2}}{1+\beta_1/\beta_2} \times \sqrt{\frac{1}{(1+1/b_1)(1+1/b_2)}},
\end{aligned}$$

where we use the fact that the absolute timescale is immaterial. The first term achieves a global maximum of $1/2$ at $\beta_1 = \beta_2$. The second is strictly smaller than 1, but asymptotically approaches 1 as b_1, b_2 jointly approach infinity. All downstream processes are stochastic and desynchronize molecular observables. Therefore, $1/2$ is the supremum of gene-gene correlations in this class of models.

Fully correlated gene loci Conversely, we can consider the two-gene problem assuming that the burst distributions are identical and fully correlated. Physically, this model may correspond to coupling of *initiation* processes, e.g. this may occur when two genes are controlled by a single promoter. This burst distribution has the following joint PGF:

$$\begin{aligned}
F(x_1, x_2) &= \sum_{i,j} x_1^i x_2^j p_{i,j} \\
&= \sum_{i,j} x_1^i x_2^j p_i \delta(i-j) = \sum_{i=0}^{\infty} (x_1 x_2)^i \left(\frac{b}{b+1}\right)^i \frac{1}{b+1} \\
&= \frac{1}{1+b-b x_1 x_2} \\
M(u_1, u_2) &= F(1+u_1, 1+u_2) = \frac{1}{1+b-b(1+u_1)(1+u_2)};
\end{aligned}$$

upon inserting the characteristics, we yield

$$\begin{aligned}
\phi &= k\int_0^\infty \left[\frac{1}{1+b-b(1+u_1e^{-\beta_1 s})(1+u_2e^{-\beta_2 s})}-1\right]ds \\
\frac{d^2e^\phi}{du_1du_2} &= \frac{d}{du_1}e^\phi k\int_0^\infty \frac{b(1+u_1e^{-\beta_1 s})e^{-\beta_2 s}ds}{(1+b-b(1+u_1e^{-\beta_1 s})(1+u_2e^{-\beta_2 s}))^2} \\
&= e^\phi k^2\int_0^\infty \frac{b(1+u_1e^{-\beta_1 s})e^{-\beta_2 s}ds}{(1+b-b(1+u_1e^{-\beta_1 s})(1+u_2e^{-\beta_2 s}))^2} \\
&\times \int_0^\infty \frac{b(1+u_2e^{-\beta_2 s})e^{-\beta_1 s}ds}{(1+b-b(1+u_1e^{-\beta_1 s})(1+u_2e^{-\beta_2 s}))^2} \\
&- e^\phi k\int_0^\infty \frac{be^{-(\beta_1+\beta_2)s}ds}{(1+b-b(1+u_1e^{-\beta_1 s})(1+u_2e^{-\beta_2 s}))^2} \\
&+ e^\phi k\int_0^\infty \frac{2be^{-(\beta_1+\beta_2)s}(1+b)ds}{(1+b-b(1+u_1e^{-\beta_1 s})(1+u_2e^{-\beta_2 s}))^3}
\end{aligned}$$

Plugging in $u_1 = u_2 = 0$,

$$\begin{aligned}
\mathbb{E}[\Lambda_1\Lambda_2] &= k^2b^2\int_0^\infty e^{-\beta_1 s}ds\int_0^\infty e^{-\beta_2 s}ds \\
&- kb\int_0^\infty e^{-(\beta_1+\beta_2)s}ds + 2kb(1+b)\int_0^\infty e^{-(\beta_1+\beta_2)s}ds \\
&= k^2b^2\frac{1}{\beta_1}\frac{1}{\beta_2} - kb\frac{1}{\beta_1+\beta_2} + 2kb(1+b)\frac{1}{\beta_1+\beta_2} \\
&= \mu_1\mu_2 + \frac{kb(2b+1)}{\beta_1+\beta_2},
\end{aligned}$$

which yields the result

$$\mathrm{Cov}(\Lambda_1,\Lambda_2) = \frac{kb(2b+1)}{\beta_1+\beta_2}.$$

Therefore, the correlation is

$$\begin{aligned}
\rho &= \frac{\mathrm{Cov}(\Lambda_1,\Lambda_2)}{\sigma_1\sigma_2} = \frac{kb(2b+1)}{\beta_1+\beta_2}\times\sqrt{\frac{\beta_1\beta_2}{k^2b^2(1+b)^2}} \\
&= \frac{2b+1}{\beta_1+\beta_2}\times\frac{\sqrt{\beta_1\beta_2}}{1+b} \\
&= \frac{\sqrt{\beta_1\beta_2}}{\beta_1+\beta_2}\times\frac{2b+1}{b+1} = \frac{\sqrt{\beta_1\beta_2}}{\beta_1+\beta_2}\times\left(\frac{b+1}{b+1}+\frac{b}{b+1}\right) \\
&= \frac{\sqrt{\beta_1/\beta_2}}{1+\beta_1/\beta_2}\times\left(1+\frac{b}{b+1}\right).
\end{aligned}$$

As in the case of uncoupled gene sizes reported in Equation 1, the first term is at most 1/2. The second term asymptotically approaches 2 as $b\to\infty$. Therefore, there are no intrinsic model constraints on Pearson correlation coefficients of two gene products; constraints arise as the *effect* of the burst size correlation structure.

Anti-correlated gene loci With these results in mind, we can consider the problem of describing genes with high *negative* correlations. As an illustration, we can consider two genes driven by a single telegraph process: gene 1 is on whenever gene 2 is off and vice versa; therefore, their respective products $\mathcal{T}_1$ and $\mathcal{T}_2$ must have a negative correlation. However, only one of these genes has a well-defined bursty limit that is infinitesimally short on periods; the other will essentially be on all the time and transcribe constitutively, containing no mutual information about the burst timing.
Nevertheless, putting aside the problem of positing a specific limiting mechanism, we can ask whether *any* joint burst distribution can produce negative correlations in molecule counts, *despite* perfect synchronization between burst events. Considering the cross moment of mRNA produced at two synchronized loci:

$$\begin{aligned}
\phi &= k\int_0^\infty \left[M(U_1(u_1), U_2(u_2)) - 1\right] ds \\
\frac{d^2 e^\phi}{du_1 du_2} &= \frac{d}{du_1} e^\phi k \int_0^\infty \frac{dM}{du_2} ds \\
&= e^\phi k \int_0^\infty \frac{d^2 M}{du_1 u_2} ds + e^\phi k^2 \int_0^\infty \frac{dM}{du_1} ds \int_0^\infty \frac{dM}{du_2} ds \\
\mathbb{E}[\Lambda_1 \Lambda_2] &= k\int_0^\infty \frac{d^2 M}{du_1 u_2} ds + k^2 \int_0^\infty \frac{dM}{du_1} ds \int_0^\infty \frac{dM}{du_2} ds,
\end{aligned}$$

with the partial derivatives evaluated at $u_1 = u_2 = 0$. The second term matches $\mu_1\mu_2$, and is strictly positive. The first term is the integral of an exponentially discounted burst cross moment:

$$\begin{aligned}
\frac{d^2 M}{du_1 u_2} &= \frac{d}{du_2}\left(\frac{dM}{dU_1}\frac{dU_1}{du_1}\right) \\
&= \frac{d^2 M}{dU_1 dU_2}\frac{dU_1}{du_1}\frac{dU_2}{du_2} \\
&= \frac{d^2 M}{dU_1 dU_2} e^{-(\beta_1+\beta_2)s} \\
&= \mathbb{E}[B_1 B_2] e^{-(\beta_1+\beta_2)s}
\end{aligned}$$

where the partial derivatives are yet again evaluated at $u_1 = u_2 = U_1 = U_2 = 0$, and B_1 and B_2 denote the SDE jump sizes at the two gene loci. By the definition of covariance:

$$\mathbb{E}[\Lambda_1 \Lambda_2] = k\frac{\mathbb{E}[B_1 B_2]}{\beta_1+\beta_2} + \mu_1\mu_2 = \frac{k}{\beta_1+\beta_2}(\mathrm{Cov}(B_1, B_2) + \mu_{B_1}\mu_{B_2}) + \mu_1\mu_2.$$

Now, supposing the correlation between the burst sizes is $\rho \in [-1, 0)$, and considering the covariance between the transcripts:

$$\mathrm{Cov}(\Lambda_1, \Lambda_2) = k\frac{\mathbb{E}[B_1 B_2]}{\beta_1+\beta_2} = \frac{k}{\beta_1+\beta_2}(\rho\sigma_{B_1}\sigma_{B_2} + \mu_{B_1}\mu_{B_2}),$$

which achieves a minimum at $\rho = -1$. Thus, the covariance has a lower limit:

$$\mathrm{Cov}(\Lambda_1, \Lambda_2) \geq \frac{k}{\beta_1+\beta_2}(\mu_{B_1}\mu_{B_2} - \sigma_{B_1}\sigma_{B_2}).$$

Without constructing the joint distribution explicitly, if we suppose the marginal discrete burst distributions are geometric – i.e., the jump sizes are exponential – then $\mu_{B_i} = \sigma_{B_i} = b_i$, and the lower limit on covariance is zero. This means that negative correlations cannot possibly result from a model with geometrically distributed, synchronized jumps. However, other joint burst laws *can* produce negative correlations, as long as the population correlation coefficient is sufficiently negative and the burst distributions are sufficiently dispersed.

We can demonstrate the existence of processes with negative count correlations induced by synchronized burst events. First, we suppose that the marginal burst distributions are identical and described by a gamma law with shape α and scale θ, enforcing $\mu_{B_1} = \mu_{B_2} = \alpha\theta$ and $\sigma^2_{B_1} = \sigma^2_{B_2} = \alpha\theta^2$. Therefore, the covariance of the Poisson intensities takes the following form:

$$\begin{aligned}\mathrm{Cov}(\Lambda_1, \Lambda_2) &= \frac{k}{\beta_1 + \beta_2}(\rho\alpha\theta^2 + \alpha^2\theta^2) \\ &= \frac{k\theta^2}{\beta_1 + \beta_2}(\rho\alpha + \alpha^2),\end{aligned}$$

which achieves $\mathrm{Cov}(\Lambda_1, \Lambda_2) < 0$ whenever $\rho\alpha + \alpha^2 < 0$. Therefore, for any $\rho \in (-1, 0)$, every $\alpha \in (0, -\rho)$ meets this criterion.

It remains to confirm that a bivariate gamma distribution with a negative correlation can exist. Such a distribution was constructed by Moran, and permits all $\rho \in (-1, 1)$ [29,30]. Furthermore, a simple application of the Cauchy-Schwartz inequality yields [31]:

$$M(u_1, u_2) = \mathbb{E}[e^{u_1 B_1 + u_2 B_2}] \leq \sqrt{\mathbb{E}[e^{2u_1 B_1}]\mathbb{E}[e^{2u_2 B_2}]} < \infty;$$

therefore, the joint MGF of the correlated bivariate gamma distribution is guaranteed to exist. This demonstrates the existence of continuous moving average processes with negative stationary correlation, driven by one Poisson process arrival process. Finally, the corresponding Poisson mixture has identical covariance, and must also have a negative correlation. Therefore, a CME with marginal negative binomial burst distributions and a carefully chosen joint structure can achieve negative molecular correlations, even if the bursts are synchronized.

Multi-gene dynamics emerging from fast processing Interestingly, there is a set of single-gene systems that recapitulate the multi-gene functional form in the limit of fast splicing. Consider a source species $\mathcal{T}_0$, which is produced in geometrically distributed bursts and converted to species $\{\mathcal{T}_i\}$, with $i = 1, ..., N$, at rates β_i. These transcripts are degraded at rates γ_i.

Furthermore, suppose all of the $\beta_i \sim \mathcal{O}(\varepsilon^{-1})$ for small ε, i.e., the source transcript is extremely unstable. In this limit, $\mathcal{T}_i$ are produced with bursts of size $B\beta_i/r$, where B is the underlying $\mathcal{T}_0$ burst size, $r := \sum_i \beta_i$, and the ratio β_i/r is $O(1)$. We define corresponding *weights* $w_i := \beta_i/r$; by definition, $\sum_i w_i = 1$. This yields:

$$
\begin{aligned}
U_0 &= Ke^{-rs} + \sum_{i=1}^{N} \frac{\beta_i u_i}{r-\gamma_i} e^{-\gamma_i s} \\
&\to \sum_{i=1}^{N} w_i u_i e^{-\gamma_i s} \\
M(U_0) &= \frac{1}{1-bU_0} = \frac{1}{1-b\sum_i w_i u_i e^{-\gamma_i s}},
\end{aligned}
$$

where M is recognizable as the joint MGF of a perfectly correlated N-variate exponential distribution with marginal distributions $B_i = w_i B$:

$$
M(u_1, ..., u_N) = \mathbb{E}\left[e^{\sum_i u_i B_i}\right] = \mathbb{E}\left[e^{(\sum_i w_i u_i)B}\right],
$$

i.e., the univariate exponential MGF evaluated at $\sum_{i=1}^{N} w_i u_i$.
The corresponding discrete PGF is:

$$
\begin{aligned}
F(x_1, ..., x_N) &= \frac{1}{1-b\sum_i w_i(x_i-1)} = \frac{1}{1-b\sum_i (w_i x_i - w_i)} \\
&= \frac{1}{1+b-b\sum_i w_i x_i}
\end{aligned}
$$

As seen above, this is *not* the perfectly correlated multivariate geometric distribution: the stochasticity of the reaction channel selection is non-negligible. Instead, we can construct a distribution over $\{0,1\}^N$, with the probability of state δ_{ij} (i.e., the vector contains a one at position i) set to w_i. It is easy to see that the generating function of the random variable $Z \in \{0,1\}^N$ takes the form derived above:

$$
H(x_1, ..., x_N) = \mathbb{E}\left[x_1^{Z_1} ... x_N^{Z_N}\right] = \sum_i w_i x_i,
$$

which amounts to $F(x_1, ..., x_N) = F(H(x_1, ..., x_N))$; i.e., the effective gene dynamics are described by a compound distribution.
Upon inserting the characteristics and selecting a set of two genes (arbitrarily indexed by 1 and 2), we yield:

$$
\begin{aligned}
\phi &= k\int_0^\infty \left[\frac{1}{1-bw_1u_1e^{-\gamma_1 s} - bw_2u_2e^{-\gamma_2 s}} - 1\right] ds \\
\frac{d^2e^\phi}{du_1du_2} &= \frac{d}{du_1} e^\phi k \int_0^\infty \frac{bw_2e^{-\gamma_2 s}ds}{(1-bw_1u_1e^{-\gamma_1 s} - bw_2u_2e^{-\gamma_2 s})^2} \\
&= e^\phi k^2 \int_0^\infty \frac{bw_1e^{-\gamma_1 s}ds}{(1-bw_1u_1e^{-\gamma_1 s} - bw_2u_2e^{-\gamma_2 s})^2} \\
&\quad \times \int_0^\infty \frac{bw_2e^{-\gamma_2 s}ds}{(1-bw_1u_1e^{-\gamma_1 s} - bw_2u_2e^{-\gamma_2 s})^2} \\
&\quad + e^\phi k \int_0^\infty \frac{2b^2w_1w_2e^{-(\gamma_1+\gamma_2)s}ds}{(1-bw_1u_1e^{-\gamma_1 s} - bw_2u_2e^{-\gamma_2 s})^3}.
\end{aligned}
$$

Plugging in $u_1 = u_2 = 0$,

$$\begin{aligned}\mathbb{E}[\Lambda_1\Lambda_2] &= k^2b^2w_1w_2\int_0^\infty e^{-\gamma_1 s}ds\int_0^\infty e^{-\gamma_2 s}ds + 2kb^2w_1w_2\int_0^\infty e^{-(\gamma_1+\gamma_2)s}ds \\ &= k^2b^2\frac{w_1}{\gamma_1}\frac{w_2}{\gamma_2} + 2kb^2\frac{w_1w_2}{\gamma_1+\gamma_2} \\ &= \mu_1\mu_2 + \frac{2kb^2w_1w_2}{\gamma_1+\gamma_2},\end{aligned}$$

since marginalization with respect to any gene recovers the univariate geometric burst distribution with scale bw_i, which immediately yields $\mu_i = kbw_i/\gamma_i$. This implies:

$$\mathrm{Cov}(\Lambda_1,\Lambda_2) = \frac{2kb^2w_1w_2}{\gamma_1+\gamma_2}.$$

From the marginal results, we still have $\sigma_i^2 = kbw_i(1+bw_i)/\gamma_i$. This yields:

$$\begin{aligned}\rho = \frac{\mathrm{Cov}(\Lambda_1,\Lambda_2)}{\sigma_1\sigma_2} &= \frac{2kb^2w_1w_2}{\gamma_1+\gamma_2}\times\sqrt{\frac{\gamma_1\gamma_2}{k^2bw_1(1+bw_1)bw_2(1+bw_2)}} \\ &= \frac{2\sqrt{\gamma_1\gamma_2}}{\gamma_1+\gamma_2}\times\sqrt{\frac{b^2w_1^2w_2^2}{w_1w_2(1+bw_1)(1+bw_2)}} \\ &= \frac{2\sqrt{\gamma_1/\gamma_2}}{1+\gamma_1/\gamma_2}\times\sqrt{\frac{1}{(1+\frac{1}{bw_1})(1+\frac{1}{bw_2})}} \in (0,1).\end{aligned}$$

Therefore, fast processing of the source transcript yields dynamics equivalent to a class of multi-gene models, with positive, but otherwise unconstrained, correlation between the downstream species.

2.7 Transient gene dynamics

Thus far, we have primarily focused on stationary systems with time-independent parameters. Nevertheless, there are classes of physiological phenomena, such as differentiation and cell cycling, where transient behaviors are crucial, particularly since these processes occur on timescales comparable to the mRNA lifetimes [1,32]. Usually, the regulatory events underpinning these processes are modeled by variation in DNA-localized transcriptional parameters [33–36].

By examination of the generating function relations, it is easy to see that the current framework is trivial to extend to *any* deterministic variation in k and M:

$$\phi(t) = \int_0^t k(s)[M(\mathbf{U}(s),s)-1]ds,$$

where we have adopted the shorthand $\mathbf{U}$ for the set of exponential sums $U_1(s),...,U_N(s)$ governing each burst product of the N loci. Therefore, burst frequency, burst size, and even the number of synchronized gene loci per cell can vary, continuously or discontinuously.

If the reaction rates within $\mathbf{U}$ change over time, the generating function PDE then becomes intractable for all but the simplest models, such as piecewise constant. The gene locus parameters b and k can also vary stochastically. In principle, certain models can be solved by defining an appropriate SDE-CME system. However, such dynamics are generally challenging to treat without recourse to numerical ODE solvers. Further, we restrict our analysis to systems tractable *via* the PGF, which cannot be applied to continuous-valued parameters.

3 Applications to sequencing data

The current framework provides quadrature-based solutions to the forward problem of PMF prediction for a broad set of transcriptional processes. More broadly, we would like to treat the *inverse* problem of identifying parameters from sequencing data. A wide range of statistical approaches are available; however, in practice, even the simplest, ergodic version of the the inverse problem depends on the following prerequisites:

1. Single-cell, single-molecule data for a set of cells in local equilibrium. This information permits the application of the ergodic model.

2. Full annotation of intermediate transcripts, *including causal relationships*, such as the splicing graph and the identities of degraded molecules. This information permits mapping between experimental data and theoretical quantities.

3. Transcriptome-wide quantification of all transcripts, ideally unbiased and fully saturated – or, at the very least, imperfect quantification combined with a statistical model of sequencing. This information permits the inclusion of technical noise.

No perfect experimental solution exists. The collection of single-cell, single-molecule data is enabled by barcoding [37]. Characterization of splicing graphs has been treated in experimental [38,39] and computational [40] contexts. However, these necessarily rely on comprehensive single-molecule annotations – which distinct intron/exon combinations occur? – which have only become feasible on a genome-wide scale since the introduction of molecular barcoding. Fully saturated sequencing is infeasible due to cost, and potentially due to thermodynamic constraints. Finally, standard sequencing capture protocols are, by design, biased toward polyadenylated regions [37]. This effect has been exploited to capture natural intronic sequences [1] and synthetic antibody-conjugated oligonucleotides [41,42], but the quantitative effects of these biases are as of yet unclear. Naturally, these data quality challenges are compounded with standard statistical challenges, such as the often considerable computational expense of determining confidence regions.

In spite of these challenges, we *can* immediately apply some of the simpler theoretical results to transcriptomic datasets.

We used long-read sequencing data generated by FLT-seq (full-length transcript sequencing by sampling) [43]. In brief, the method synthesizes a cDNA library using 10X gel beads and primers, amplifies it, then applies nanopore sequencing to generate long reads with associated cell and molecule barcodes. We obtained a mouse stem cell dataset (498 barcodes), filtered for the activated subset (136 barcodes), and identified the 1000 highest-expressed genes.

We used *gffutils* 0.10.1 to construct a database of identified intermediate and terminal isoforms, based on the accompanying annotations generated by the computational pipeline *FLAMES* (full-length analysis of mutations and splicing). We used the *intervaltree* 2.1.0 Python package to split the transcript-specific exons into elementary intervals, then used graph tools from the *NetworkX* 2.5.1 Python package to identify "root" transcripts that cannot be obtained by removing regions of any other transcript.

The presumed source transcript covering the entire locus was not observed for any of the genes. As their exonic patterns are mutually exclusive, we model each gene's root transcripts as products of a single rapidly processed source species, as in Section 2.6.3. The theoretical results immediately imply that the root transcripts must be distributed according to the negative binomial distribution.

Therefore, by fitting the transcript distributions, we can estimate effective burst sizes bw_i and unitless efflux rates γ_i/k. These marginal parameters can be plugged into the formula for Pearson correlation derived in Section 2.6.3, and compared to the empirical correlation coefficients. The theoretical correlations are to be understood as upper limits, as unobserved intermediate processing steps and technical noise inevitably reduce transcript–transcript correlations.

We fit the root transcript marginal distributions using the *statsmodels* 0.10.2 Python package. One gene was rejected outright due to failure to construct elementary intervals. 732 transcripts were rejected due to underdispersion (mean higher than variance), as they do not possess valid maximum likelihood estimates. 1290 transcripts were rejected due to poor fit (failure to converge or excessively sparse data, with all but one rejected transcripts having average expression below 1 mRNA per cell). This analysis produced a set of 4362 transcripts and 12480 nontrivial correlation matrix entries.

Theoretical and observed correlation coefficients are shown in Figure 6. 11894 of the theoretical correlations (95.3%) are higher than the observed ones, whereas 586 (4.7%) are lower. Furthermore, the theoretical correlations are clearly nontrivial; all are less than unity. The results suggest that the model is not sufficient to recapitulate the full dynamics, but *does* provide an effective theoretical constraint. We anticipate that the lower observed correlations emerge from a contribution of technical noise in the sequencing process, suboptimal inference from the marginals, correlation-degrading stochastic intermediates, and model misidentification. The final effect is particularly plausible for the cluster evident on the left side of the figure, as the fast-processing model cannot account for negative transcript–transcript correlations.

4 Results and discussion

We have described a broad extension of previous work pertaining to monomolecular reaction networks coupled to a bursty transcriptional process. In particular, by exploiting the standard properties of reaction rate equations, we have demonstrated the existence of all moments and cross-moments. Further, we have derived the analytical expressions for the generating functions and demonstrated their existence. The following expression gives the general solution for the factorial-cumulant generating function:

$$\phi(t) = \int_0^t k(s)[M(\mathbf{U}(s), s) - 1]ds, \tag{2}$$

where $\mathbf{U}$ is a set of functions associated with the gene products, computable by the procedure in Section 2.2.3, and M is a joint generating function governing the bursting dynamics.

Under the mild assumptions of finite-activity bursting and Markovian downstream processes, the expressions hold for arbitrary directed acyclic graphs of splicing and degradation, coupled to an arbitrary number of gene loci with arbitrary, and potentially correlated and time-dependent, burst dynamics. In fact, we explicitly consider the problem of modeling multiple synchronized genes and find that geometric burst size coordination is *required* to achieve transcript count correlations $\rho > 1/2$. Furthermore, we test whether negative correlations are feasible under the assumption of synchronized bursts at multiple gene loci, and find that $\rho < 0$ are impossible with geometric bursts, but *can* be achieved with negative binomial bursts. These results substantially constrain and inform the space of models that can recapitulate the combination of bursty dynamics [11] and high absolute gene-gene correlations [3, 27] observed in living cells.

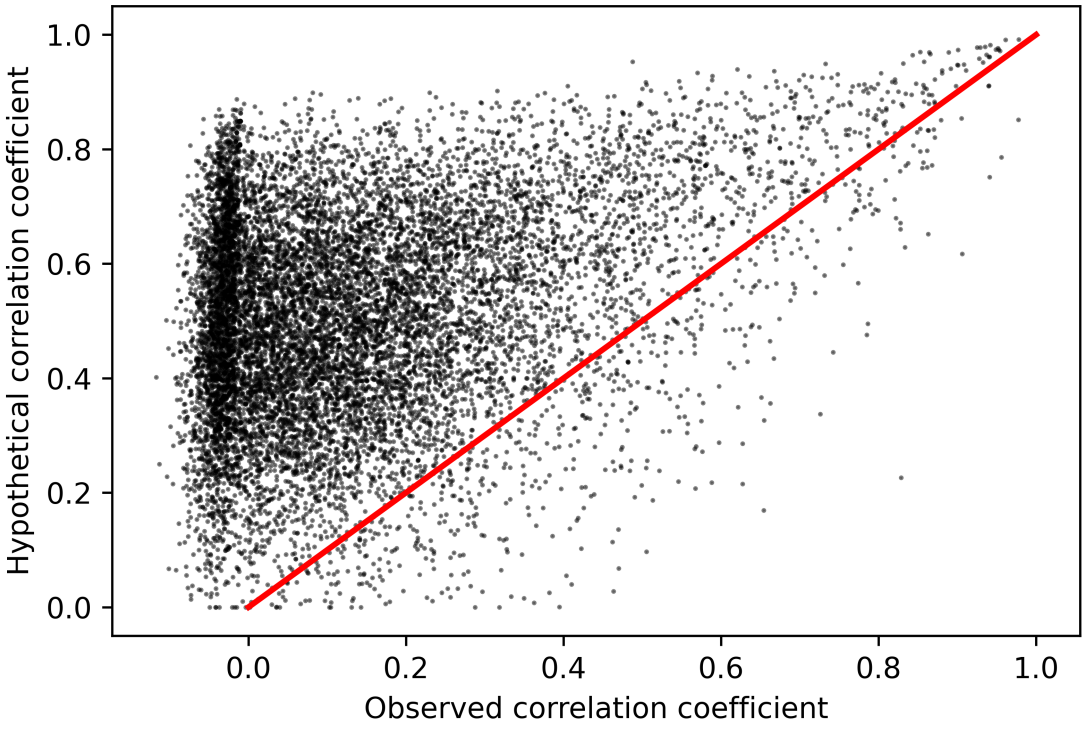

Figure 6: Theoretical Pearson correlation coefficients inferred from marginal distributions, compared to empirical correlations. The theoretical model generally overestimates transcript–transcript correlations, and rarely underestimates them, which is consistent with the interpretation of the theoretical correlation as a nontrivial upper limit.

We compared the theoretical constraints with experimental data generated by FLT-seq, a recent long-read, single-cell, single-molecule sequencing method. Investigating a set of 4362 transcripts, we found that the constraints were met for 95.3% of the transcript–transcript correlations. Nevertheless, the model was insufficient to recapitulate the precise quantitative details, suggesting that more detailed biophysical models of splicing networks and technical noise are necessary.

Although we primarily focus on distributions, with an eye to inference from atemporal data, the solution is robust in several complementary dimensions. Firstly, the upper limit of the integral in Equation 2 is arbitrary, and can be evaluated up to a finite time horizon to yield a transient distribution. Secondly, the formulation of the problem presupposes $m_i(t=0)=0$ for all species i. However, if nonzero molecule counts are present at $t=0$, it is straightforward to compute the resulting log-PGF via $\phi = \phi^h + \sum_{i=1}^{n} m_i(t=0)\ln(1+U_i)$, where ϕ^h is the result for $m_i(t=0)=0$ and the U_i are species-specific exponential sums. This relation produces arbitrary *conditional* distributions, as derived elsewhere [44]. Therefore, the current approach can be used to compute the likelihoods of entire trajectories of observations, and thus perform parameter estimation on live-cell data.

The computational complexity of this procedure is $O(\mathcal{N}\log\mathcal{N})$ in state space size. However, this

complexity ostensibly arises from the n-dimensional inverse Fourier transform. We expect the time complexity to be $O(\mathcal{N})$ in the practical regime, and irrelevant outside due to the *space* complexity of holding an array of size $\mathcal{N}$ in memory for the inverse Fourier transform.

Curiously, this class of analytical solutions to reaction networks can be adapted to a subset of diffusion problems. General diffusion on a multidimensional lattice is not directly solvable, because it violates the assumption of acyclic graph structure. However, percolation through a directed acyclic graph coupled to a source and a set of sinks can be described using the current mathematical formalism. Further, such a percolation can represent the incremental movement of RNA polymerase along a DNA strand, integrating discrete copy number statistics with submolecular details in a single analytical framework [45, 46].

We do not treat the aforementioned auxiliary degenerate solutions that arise when $\beta_i = \beta_j$ for some i, j. However, Section 2.4 guarantees the existence of these solutions and all moments: the procedure does not rely on a particular functional form of ψ_i, only its existence, which is implied by the test subordinator $F(x) = x$ matching the constitutive case [14].

Finally, we briefly touch upon the class of delay chemical master equations, and survey several recent advances in the field in Section S1.1. Due to the non-Markovian nature of delayed systems, general probabilistic solutions are rare [47] and represent an open area of study. In our discussion, we motivate delays as a limit of numerous, fast isomerization processes, and clarify the challenges inherent in applying the analysis of delay CMEs to bursty systems.

5 Appendix: ODE solution

The solution to the generic ODE $\frac{dy}{dx} = \sum_i a_i e^{-b_i x} - ry$ is $y = \sum_i \frac{a_i}{r-b_i} e^{-b_i x} + Ke^{-rx}$. Using the initial condition $y(0) = \xi$ yields $K = \xi - \sum_i \frac{a_i}{r-b_i} a_i$. Consider an efflux rate $r = \sum_j \beta_j + \gamma$, i.e., a set of isomerization pathways and a single degradation pathway from the species in question. Care must be taken when the downstream paths converge (as in the case of the two-intron system): duplicate terms in product characteristics U_j need to be aggregated, with $\sum_j \beta_j U_j$ rewritten as $\sum_i a_i e^{-b_i x}$. This is computationally straightforward to do by choosing appropriate data structures, as with the matrix A in the implementation of the DAG algorithm.

6 Code availability

Python notebooks that can be used to reproduce Figures 5 and 6 are available at `https://github.com/pachterlab/GP_2021_2`.

7 Acknowledgments

The DNA illustration used in Figures 1 and 4, modified from [44], is a derivative of the DNA Twemoji by Twitter, Inc., used under CC-BY 4.0. The directed acyclic graph generation code was adapted from the IPython Parallel reference documentation: `https://ipyparallel.readthedocs.io/en/latest/dag_dependencies.html`. The yellow and gray colors used in Figures 1, 3, and 4 are the Pantone Colors of the Year 2021, PANTONE 17-5104 Ultimate Gray and PANTONE 13-0647 Illuminating. G.G. and L.P. were partially funded by NIH U19MH114830.

References

[1] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, August 2018.

[2] Sheel Shah, Yodai Takei, Wen Zhou, Eric Lubeck, Jina Yun, Chee-Huat Linus Eng, Noushin Koulena, Christopher Cronin, Christoph Karp, Eric J. Liaw, Mina Amin, and Long Cai. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell*, 174(2):363–376.e16, July 2018.

[3] Alex K. Shalek, Rahul Satija, Xian Adiconis, Rona S. Gertner, Jellert T. Gaublomme, Raktima Raychowdhury, Schraga Schwartz, Nir Yosef, Christine Malboeuf, Diana Lu, John J. Trombetta, Dave Gennert, Andreas Gnirke, Alon Goren, Nir Hacohen, Joshua Z. Levin, Hongkun Park, and Aviv Regev. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, 498(7453):236–240, June 2013.

[4] Carlos F Buen Abad Najar, Nir Yosef, and Liana F Lareau. Coverage-dependent bias creates the appearance of binary splicing in single cells. *eLife*, 9:e54603, June 2020.

[5] Brian Munsky, Brooke Trinh, and Mustafa Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5, 2009.

[6] Brian Munsky, Zachary Fox, and Gregor Neuert. Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics. *Methods*, 85:12–21, 2015.

[7] Gennady Gorin and Lior Pachter. Intrinsic and extrinsic noise are distinguishable in a synthesis – export – degradation model of mRNA production. Preprint, bioRxiv: 10.1101/2020.09.25.312868, September 2020.

[8] M. Ullah and O. Wolkenhauer. Family tree of Markov models in systems biology. *IET Systems Biology*, 1(4):247–254, July 2007.

[9] Crispin Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer, 3 edition, 2004.

[10] Abhyudai Singh and Pavol Bokes. Consequences of mRNA Transport on Stochastic Variability in Protein Levels. *Biophysical Journal*, 103(5):1087–1096, September 2012.

[11] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences*, 109(43):17454–17459, October 2012.

[12] Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell*, 123(6):1025–1036, December 2005.

[13] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, December 1976.

[14] Tobias Jahnke and Wilhelm Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1):1–26, December 2006.

[15] Lisa Amrhein, Kumar Harsha, and Christiane Fuchs. A mechanistic model for the negative binomial distribution of single-cell mRNA counts. Preprint, bioRxiv: 657619, June 2019.

[16] C. W. Gardiner and S. Chaturvedi. The poisson representation. I. A new technique for chemical master equations. *Journal of Statistical Physics*, 17(6):429–468, December 1977.

[17] Chris Chatfield, Anne B. Koehler, J. K. Ord, and Ralph D. Snyder. A New Look at Models For Exponential Smoothing. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50(2):147–159, July 2001.

[18] Harry H Panjer. Mixed Poisson Distributions. In *Encyclopedia of Actuarial Science*. John Wiley & Sons, Ltd, 2004.

[19] Nir Friedman, Long Cai, and X. Sunney Xie. Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression. *Physical Review Letters*, 97(16):168302, October 2006.

[20] Rama Cont and Peter Tankov. *Financial Modeling with Jump Processes*. Financial Mathematics. Chapman & Hall, 2004.

[21] Nicola Cufaro Petroni and Piergiacomo Sabino. Gamma Related Ornstein-Uhlenbeck Processes and their Simulation. Preprint, arXiv: 2003.08810, March 2020.

[22] Ole E. Barndorff-Nielsen and Neil Shephard. Integrated OU Processes and Non-Gaussian OU-based Stochastic Volatility Models. *Scandinavian Journal of Statistics*, 30(2):277–295, June 2003.

[23] Ole E Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in Financial economics. *Journal of the Royal Statistical Society: Series B*, 63:167–241, 2001.

[24] Lucy Ham, Rowan D. Brackston, and Michael P.H. Stumpf. Extrinsic Noise and Heavy-Tailed Laws in Gene Expression. *Physical Review Letters*, 124(10):108101, March 2020.

[25] Ole. E. Barndorff-Nielsen and Steen Thorbjørnsen. Self-Decomposability and Lévy Processes in Free Probability. *Bernoulli*, 8(3):323–366, 2002. Publisher: International Statistical Institute (ISI) and Bernoulli Society for Mathematical Statistics and Probability.

[26] Dimitris Karlis and Evdokia Xekalaki. Mixed Poisson Distributions. *International Statistical Review / Revue Internationale de Statistique*, 73(1):35–58, 2005. Publisher: [Wiley, International Statistical Institute (ISI)].

[27] Giovanni Iacono, Ramon Massoni-Badosa, and Holger Heyn. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biology*, 20(1):110, December 2019.

[28] Heng Xu, Leonardo A Sepúlveda, Lauren Figard, Anna Marie Sokac, and Ido Golding. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nature Methods*, 12(8):739–742, August 2015.

[29] P.A.P Moran. Statistical inference with bivariate gamma distributions. *Biometrika*, 56(3):627–634, 1969.

[30] S. Yue, T.B.M.J. Ouarda, and B. Bobée. A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, 246(1-4):1–18, June 2001.

[31] Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. Texts in Statistical Science. CRC Press, Taylor & Francis Group, 2015.

[32] Ron Milo and Rob Phillips. *Cell Biology by the Numbers*. Garland Science, July 2015.

[33] Samuel O Skinner, Heng Xu, Sonal Nagarkar-Jaiswal, Pablo R Freire, Thomas P Zwaka, and Ido Golding. Single-cell analysis of transcription kinetics across the cell cycle. *eLife*, 5:e12175, January 2016.

[34] Justine Dattani and Mauricio Barahona. Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *Journal of The Royal Society Interface*, 14(126):20160833, January 2017.

[35] James A. Briggs, Caleb Weinreb, Daniel E. Wagner, Sean Megason, Leonid Peshkin, Marc W. Kirschner, and Allon M. Klein. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, 360(6392):eaar5780, June 2018.

[36] A. Zeisel, W. J. Kostler, N. Molotski, J. M. Tsai, R. Krauthgamer, J. Jacob-Hirsch, G. Rechavi, Y. Soen, S. Jung, Y. Yarden, and E. Domany. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular Systems Biology*, 7(1):529–529, April 2014.

[37] Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, Mark T. Gregory, Joe Shuga, Luz Montesclaros, Jason G. Underwood, Donald A. Masquelier, Stefanie Y. Nishimura, Michael Schnall-Levin, Paul W. Wyatt, Christopher M. Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D. Ness, Lan W. Beppu, H. Joachim Deeg, Christopher McFarland, Keith R. Loeb, William J. Valente, Nolan G. Ericson, Emily A. Stevens, Jerald P. Radich, Tarjei S. Mikkelsen, Benjamin J. Hindson, and Jason H. Bielas. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, April 2017.

[38] O Kessler, Y Jiang, and L A Chasin. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Molecular and Cellular Biology*, 13(10):6211–6222, October 1993.

[39] Heather L. Drexler, Karine Choquet, and L. Stirling Churchman. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Molecular Cell*, 77(5):985–998.e8, March 2020.

[40] Steffen Heber, Max Alekseyev, Sing-Hoi Sze, Haixu Tang, and Pavel A. Pevzner. Splicing graphs and EST assembly problem. *Bioinformatics*, 18(suppl_1):S181–S188, July 2002.

[41] Vanessa M Peterson, Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C Wilson, Renee Moore, Terrill K McClanahan, Svetlana Sadekova, and Joel A Klappenbach. Multiplexed quantification of proteins and transcripts in single cells. *Nature Biotechnology*, 35(10):936–939, October 2017.

[42] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14(9):865–868, September 2017.

[43] Luyi Tian, Jafar S. Jabbari, Rachel Thijssen, Quentin Gouil, Shanika L. Amarasinghe, Hasaru Kariyawasam, Shian Su, Xueyi Dong, Charity W. Law, Alexis Lucattini, Jin D. Chung, Timur Naim, Audrey Chan, Chi Hai Ly, Gordon S. Lynch, James G. Ryall, Casey J.A. Anttila, Hongke Peng, Mary Ann Anderson, Andrew W. Roberts, David C.S. Huang, Michael B. Clark, and Matthew E. Ritchie. Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing. preprint, Genomics, August 2020.

[44] Gennady Gorin and Lior Pachter. Special function methods for bursty models of transcription. *Physical Review E*, 102(2):022409, August 2020.

[45] Heng Xu, Samuel O. Skinner, Anna Marie Sokac, and Ido Golding. Stochastic Kinetics of Nascent RNA. *Physical Review Letters*, 117(12):128101, 2016.

[46] Gennady Gorin, Mengyu Wang, Ido Golding, and Heng Xu. Stochastic simulation and statistical inference platform for visualization and estimation of transcriptional kinetics. *PLOS ONE*, 15(3):e0230736, March 2020.

[47] Andre Leier and Tatiana T. Marquez-Lago. Delay chemical master equation: direct and closed-form solutions. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150049, July 2015.

[48] Norman MacDonald. *Time Lags in Biological Models*, volume 27 of *Lecture Notes in Biomathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1978.

[49] Kevin Burrage, Pamela Burrage, Andre Leier, and Tatiana T. Marquez-Lago. A Review of Stochastic and Delay Simulation Approaches in Both Time and Space in Computational Cell Biology. In David Holcman, editor, *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*. Springer International Publishing, Cham, 2017.

[50] Tomáš Gedeon and Pavol Bokes. Delayed Protein Synthesis Reduces the Correlation between mRNA and Protein Fluctuations. *Biophysical Journal*, 103(3):377–385, August 2012.

[51] Jacek Miekisz, Jan Poleszczuk, Marek Bodnar, and Urszula Foryś. Stochastic Models of Gene Expression with Delayed Degradation. *Bulletin of Mathematical Biology*, 73(9):2231–2247, September 2011.

[52] Farzad Fatehi, Yuliya N. Kyrychko, and Konstantin B. Blyuss. A new approach to simulating stochastic delayed systems. *Mathematical Biosciences*, 322:108327, April 2020.

[53] Manuel Barrio, Kevin Burrage, Andre Leier, and Tianhai Tian. Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modelling and Simulation. *PLOS Computational Biology*, 2(9):e117, September 2006.

[54] L F Lafuerza and R Toral. Exact solution of a stochastic protein dynamics model with delayed degradation. *Physical Review E*, 84:051121, November 2011.

[55] L F Lafuerza and R Toral. Role of delay in the stochastic creation process. *Physical Review E*, 84:021128, August 2011.

[56] Tao Jia and Rahul V Kulkarni. Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting. *Physical Review Letters*, 106:058102, February 2011.

S1 Supplementary Note

S1.1 Delay chemical master equations

In the current supplement, we detour from the Markovian framework to consider *delay* systems, which have deterministic, rather than stochastic, state transitions. Certain degenerate cases – for example, the problem of incremental, linear movement with identical transition rates – directly bear upon the class of delay chemical master equations (DCMEs). As an example, we can model the simple linear chain of reactions with constitutive production

$$\varnothing \xrightarrow{k} \mathcal{T}_1 \xrightarrow{\beta/n} ... \xrightarrow{\beta/n} \mathcal{T}_n \xrightarrow{\beta/n} \varnothing,$$

the total delay between production of $\mathcal{T}_1$ and degradation of $\mathcal{T}_n$ is Erlang-distributed, with shape n and rate β. As $n \to \infty$, the Erlang distribution reduces to a point mass at $\beta^{-1} := \tau$. This implies that we can treat an *aggregated* species $\mathcal{T}$, produced at rate k and degraded after a deterministic delay τ:

$$\varnothing \xrightarrow{k} \mathcal{T} \xRightarrow{\tau} \varnothing$$

This is precisely the "linear chain trick" introduced by MacDonald in 1978 [48]. The study of delayed dynamical systems, such as delay differential equations, dates back to the eighteenth century [49], with cornerstone biological models by Lotka and Volterra [48, 49]. Recent work has focused on developing exact solutions [47, 50, 51] and simulation methods [52, 53]. In particular, studies by Lafuerza and Toral [54, 55] report full analytical solutions for constitutive systems with isomerization, while a contemporary study by Jia and Kulkarni [56] reports lower moments for a system with bursty mRNA production and catalysis.

Unfortunately, applying these methods to bursty systems is challenging, and all but the simplest problems are intractable. As an illustration, we consider the constitutive two-stage system described by by Lafuerza and Toral [54], and discuss the challenges of extending it to include bursty production. If we assume that no stochastic degradation reactions occur, the reaction equations and generating function relations take the following form:

$$\varnothing \xrightarrow{k} \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \xRightarrow{\tau} \varnothing$$

$$\frac{\partial G}{\partial t} = k(F(x_1) - 1)G + \beta(x_2 - x_1)\frac{\partial G}{\partial x_1} + \beta(1 - x_2)\sum_{m_1=0}^{\infty} G^*(x_1, x_2, \tau)m_1 P(m_1, t - \tau),$$

where G^* is a conditional generating function for an auxiliary *non-degrading* system, initialized at $m_1 - 1$ molecules of the parent transcript $\mathcal{T}_1$. This auxiliary system has no degradation reactions, and allows us to incorporate the non-Markovian effects of delays. Assuming constitutive production, and using the shifted variables u_i for convenience, we find:

$$\frac{\partial G}{\partial t} = ku_1 G + \beta(u_2 - u_1)\frac{\partial G}{\partial u_1} - \beta u_2 \sum_{m_1=0}^{\infty} G^*(u_1, u_2, \tau)m_1 P(m_1, t - \tau)$$

The final term is *not* proportional to G, so no convenient exponential *ansatz* is available. However, the sum affords an alternative representation, which exploits the separability of the initial condition and the dynamics on $[0, \tau]$:

$$G^*(u_1, u_2, \tau) = [1 + U_1(\tau)]^{m_1 - 1} e^{\phi^*(\tau)}$$
$$\sum_{m_1=0}^{\infty} G^*(u_1, u_2, \tau) m_1 P(m_1, t-\tau) = e^{\phi^*(\tau)} \sum_{m_1=0}^{\infty} [1 + U_1(\tau)]^{m_1 - 1} m_1 P(m_1, t-\tau),$$

where ϕ^* is the factorial-cumulant generating function of the auxiliary system, started at zero molecules. This sum may be treated as the first derivative of the stationary $\mathcal{T}_1$ PGF, evaluated at $1 + U_1$, where U_1 is a function computed by solving the non-degrading system with the method of characteristics.

We start by computing the auxiliary U_1 by using the method of characteristics and enforcing $U_2(0) = u_2$ and $U_1(0) = u_1$.

$$\varnothing \xrightarrow{k} \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2$$
$$\frac{\partial U_2}{\partial s} = 0 \implies U_2 = u_2$$
$$\frac{\partial U_1}{\partial s} = \beta(U_2 - U_1) \implies U_1 = u_2 + (u_1 - u_2)e^{-\beta s}$$

Now, we compute the generating function of the subsystem:

$$\phi^*(t) = k \int_0^t U_1(s) ds = k \int_0^\tau \left[u_2 + (u_1 - u_2) e^{-\beta s} \right] ds$$
$$= k u_2 \tau + \frac{k}{\beta}(u_1 - u_2) \left[1 - e^{-\beta \tau} \right]$$

We compute the derivative of the $\mathcal{T}_1$ Poisson PGF:

$$H(x_1) = e^{k(x_1 - 1)/\beta}$$
$$H(u_1) = e^{k u_1 / \beta}$$
$$H'(U_1) = \frac{k}{\beta} e^{k U_1 / \beta}$$

This construction is slightly simpler than in the original: we do not use the full time-dependent Poisson distribution, but presuppose that the system starts with $\mathcal{T}_1$ in equilibrium. Since it approaches this distribution exponentially fast regardless of initial conditions, the error is minimal, and the simplification eliminates the time dependence in the degradation term.

Plugging in and evaluating the non-Markovian term:

$$e^{\phi^*(\tau)} H'(U_1(\tau)) = \frac{k}{\beta} \exp(k u_2 \tau + k u_1 / \beta)$$

Finally, considering the full generating function expression:

$$\frac{\partial G}{\partial t} = k u_1 G + \beta(u_2 - u_1) \frac{\partial G}{\partial u_1} - k u_2 e^{u_1 \frac{k}{\beta} + u_2 k \tau}$$

Lafuerza and Toral report a solution [54], though computed through an *ansatz* rather than directly – this PDE is not quite as simple as that of the Markovian system. We restrict ourselves to the stationary solution, which can be solved with a rather mechanical application of the integrating factor method, or by noticing that the uncorrelated Poisson PMF solves the equation:

$$\begin{aligned} G &= e^{u_1 \frac{k}{\beta} + u_2 k\tau} \\ \frac{\partial G}{\partial u_1} &= \frac{k}{\beta} G \\ \frac{\partial G}{\partial t} &= k u_1 G + \beta(u_2 - u_1)\frac{k}{\beta} G - k u_2 G = 0 \end{aligned}$$

Of course, this result can be derived just as well without writing down anything at all – by using the standard results for constitutive production [14], the linear chain trick, and the fact that sums of Poisson random variables are Poisson. However, the rigorous approach can be used to treat more general systems. In particular, we attempt to solve the delayed analog of the two-stage bursty system [10]:

$$\begin{aligned} \varnothing \xrightarrow{k} B \times \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \overset{\tau}{\Longrightarrow} \varnothing \\ \frac{\partial G}{\partial t} = k\left[\frac{1}{1 - b u_1} - 1\right] G + \beta(u_2 - u_1)\frac{\partial G}{\partial u_1} - \beta u_2 e^{\phi^*(\tau)} H'(U_1(\tau)), \end{aligned}$$

where the auxiliary system is now bursty.
First, we compute the factors of the non-Markovian term. The PGF derivative is found by evaluating the $\mathcal{T}_1$ marginal:

$$k\int_0^T \left[\frac{1}{1 - b u_1 e^{-\beta s}} - 1\right] ds = \frac{k}{\beta} \ln\left(\frac{b u_1 e^{-\beta T} - 1}{b u_1 - 1}\right),$$

which coincides with the relevant result for the gamma Ornstein–Uhlenbeck SDE [21]. However, this form is needlessly challenging to work with, and it is more straightforward to assume $T \gg 0$, or the system starts in the equilibrium distribution of $\mathcal{T}_1$. Again, due to exponential convergence, the error is minimal. Differentiating with respect to $x_1 = u_1 + 1$:

$$\begin{aligned} H'(x_1) &= \frac{d}{dx_1}\left(\frac{1}{1 - b(x_1 - 1)}\right)^{k/\beta} = \frac{kb}{\beta}\left(\frac{1}{1 - b(x_1 - 1)}\right)^{k/\beta + 1} \\ H'(u_1) &= \frac{\mu_1 H(u_1)}{1 - b u_1}, \end{aligned}$$

where we define $\mu_1 := kb/\beta$ for simplicity. This yields a straightforward expression for the summation:

$$\sum_{m_1=0}^{\infty} [1 + U_1]^{m_1 - 1} m_1 P(m_1, t - \tau) = \frac{\mu H(U_1)}{1 - b U_1}$$

We reuse U_1 and U_2 from the derivation of the constitutive system, as the downstream components

of the auxiliary systems match:

$$\begin{aligned}
\varnothing &\xrightarrow{k} B \times \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \\
\phi^*(\tau) &= k\int_0^\tau (M(U_1)-1)ds = k\int_0^\tau \left[\frac{1}{1-bU_1}-1\right]ds \\
&= k\int_0^\tau \left[\frac{1}{1-bu_2-b(u_1-u_2)e^{-\beta s}}-1\right]ds \\
\theta &:= \frac{b(u_1-u_2)}{1-bu_2} \\
\phi^* &= k\int_0^\tau \left[\frac{(1-bu_2)^{-1}}{1-\theta e^{-\beta s}}-1\right]ds \\
&= k\tau\left(\frac{bu_2}{1-bu_2}\right)+\frac{k}{\beta(1-bu_2)}\ln\left(\frac{\theta e^{-\beta\tau}-1}{\theta-1}\right) \\
&= k\tau\left(\frac{bu_2}{1-bu_2}\right)+\frac{k}{\beta(1-bu_2)}\ln\left(\frac{bU_1(\tau)-1}{bu_1-1}\right)
\end{aligned}$$

which follows from the derivation of the PGF of the nascent marginal.
Now, considering the full generating function relation:

$$\begin{aligned}
\frac{\partial G}{\partial t} &= k\left[\frac{1}{1-bu_1}\right]G+\beta(u_2-u_1)\frac{\partial G}{\partial u_1} \\
&\quad -\beta u_2 e^{-k\tau}\exp\left(\frac{k\tau}{1-bu_2}\right)\left(\frac{bU_1(\tau)-1}{bu_1-1}\right)^{k\beta^{-1}(1-bu_2)^{-1}}\times\frac{kb}{\beta}\left(\frac{1}{1-bU_1(\tau)}\right)^{k/\beta+1}
\end{aligned}$$

This PDE is not easily tractable by standard analytical or numerical methods. The form of the equation is rather complicated and not amenable to analysis by characteristics. In principle, a numerical PDE or ODE solver can be used: we may fix u_2 and solve for $G(u_1, u_2)$ over a mesh of u_1. By repeating this for many values of u_2, we can compute the Fourier transform of the joint distribution. However, this requires solvers that can integrate over the complex plane, as well as initial conditions $G(0, u_2)$ for each u_2. These are the very values we seek, so even numerical approaches require some ingenuity.
In short, the stochastically delayed systems reduce to deterministically delayed systems in some well-studied regimes. However, in spite of the formal connection between the CME and the DCME, the former is far simpler to analyze: the DCME is non-Markovian, and generally resistant to exact analysis. Although much recent progress has been made, regulated transcriptional systems do not yet have full probabilistic solutions.