

Modeling bursty transcription and splicing with the chemical master equation

Gennady Gorin¹ and Lior Pachter^{2,*}

¹Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California and ²Division of Biology and Biological Engineering & Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California

ABSTRACT Splicing cascades that alter gene products posttranscriptionally also affect expression dynamics. We study a class of processes and associated distributions that emerge from models of bursty promoters coupled to directed acyclic graphs of splicing. These solutions provide full time-dependent joint distributions for an arbitrary number of species with general noise behaviors and transient phenomena, 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 coexpression 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.

SIGNIFICANCE Understanding mRNA transcription requires an arsenal of tractable and versatile Markov models. Bursty transcription is ubiquitous in mammalian cells; however, only a few such systems have been mathematically characterized. We expand their scope and present full probabilistic solutions for bursty transcription at multiple, potentially synchronized, gene loci, coupled to arbitrary directed acyclic graphs of splicing. We use these general results to derive quantitative constraints on transcript count correlations, motivate physically plausible classes of burst models, and validate the constraints using single-cell sequencing data.

INTRODUCTION

Recent advances in the analysis of single-cell RNA sequencing (scRNA-seq) (1) enable the quantification of pre-mRNA molecules alongside mature mRNA. These experimental data provide an opportunity to infer the topologies and biophysical parameters governing the processes of mRNA transcription, processing, export, and degradation in living cells. In particular, they provide a novel approach to inferring and studying the dynamics of mammalian splicing cascades (2).

Faced with the enormous volume of genome-wide information accessible through this technology (3), we seek to produce models that can represent their data-generating mechanism. If the model matches the physiology, fits can be physically interpreted in terms of biological parameters, analyzed using standard Bayesian machinery, and enormously compress data sets by representing thousands of

observations by a few parameters. The choice of models is informed by a variety of considerations: physiologically plausible mechanisms based on orthogonal experiments, the form and granularity of the data, phenomenological observations that we strive to explain, and computational tractability. In this section, we use these considerations to circumscribe the scope of investigation.

We leverage the past two decades of fluorescence transcriptomics to develop consistent models. Results from live-cell profiling are consistent with Markovian dynamics; *in vivo* reactions, such as mRNA transcription and degradation, are well described by memoryless models dependent only on the instantaneous molecule counts (4,5). This also appears to be the most general class of models that affords straightforward analytical and numerical recipes; as we discuss in section S6, the complementary class of *delay* master equations, which encode molecular memory, resists systematic analysis at this time. If we do adopt the assumption of memorylessness, previous experiments immediately offer plausible models for transcriptional dynamics: mRNA transcription occurs in *bursts* of activity (6,7), with geometrically distributed bursts prevalent in bacterial and

Submitted July 30, 2021, and accepted for publication February 3, 2022.

*Correspondence: lpachter@caltech.edu

Editor: Ramon Grima.

<https://doi.org/10.1016/j.bpj.2022.02.004>

© 2022 Biophysical Society.

mammalian cells. We treat the problem in more generality, considering arbitrary burst distributions, which are mandatory for describing coexpression. Since our methods are designed for scRNA-seq, we omit all mechanistic discussion of regulation: feedback is usually modeled through *protein* interactions (8,9). Joint quantification of mRNA and proteins is in its nascence, and requires targeted antibodies, preventing genome-wide quantification (10–12). Thus, we omit stochastic regulation, although we propose that some of our results can be used to describe *deterministic* regulation without feedback.

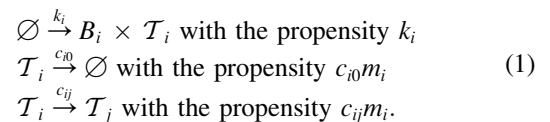
Commercially available scRNA-seq data are intrinsically discrete: they report integer counts of molecular barcodes in each cell (13). To fit these counts, we need to solve for state probabilities induced by the chemical master equation (CME), which formalizes probability flow between microstates representing copy numbers. In the most general form, the CME is an infinite series of ordinary differential equations represented by the matrix equation $\frac{dp(t)}{dt} = \mathcal{A}p(t)$, where p is a vector of probabilities and \mathcal{A} encodes the state-specific fluxes; the form of the equation, dependent only on $p(t)$ rather than any history, encodes the Markov property of memorylessness. The scRNA-seq protocols also provide base-level information, which can be used to identify specific transcripts. Inspired by the RNA velocity framework, which quantifies *unspliced* RNA along with mature, coding transcripts to infer regulatory dynamics (1), we seek to lay the groundwork for more general problems: the microstates represent joint copy-number quantities of intermediate transcripts with various introns. We do not treat the specifics of technical noise arising from the sequencing chemistry (14) and bioinformatics (15), although our solutions are modular with respect to simple noise models (16).

We would like to use physical models to encode the full scope of available data, particularly gene-gene correlations. At this time, the modeling field either uses single-gene results (treating the marginals and omitting correlations altogether) (17,18) or explicitly treats small systems (explaining correlations using concrete, usually autoregulatory schema) (19,20), whereas the sequencing field uses fully phenomenological descriptions (omitting mechanistic details of coupling) (21–23). We seek a middle ground that retains mechanistic interpretability, but offers straightforward, physically informed, and tractable recipes for representing gene-gene correlations. Finally, scRNA-seq data are particularly appealing because they can provide genome-wide information about numerous cell types and transient processes. We describe a procedure to treat random and deterministic variation in gene parameters, extending and unifying conventional models for extrinsic cell-to-cell noise (24), cell-type heterogeneity (17), and cell-cycle dynamics (25).

To enable inference, the solutions must be computationally facile. In principle, the CME can always be solved by

matrix exponentiation: the solution is simply $p(t) = p(0)e^{\mathcal{A}t}$. However, this approach does not scale very well: if the state space size is truncated to \mathcal{N} , the matrix exponential has \mathcal{N}^3 time complexity. Although reformulating the problem can ameliorate this—for example, by tensor decomposition (26,27) or treating the CME in terms of the reaction counts (28)—these approaches are poorly compatible with bursty systems, which can have an infinite number of possible reaction channels, corresponding to the infinite support of the burst distribution. More problematically, these numerical methods are not amenable to treating simpler subproblems. For example, finding lower moments or computing a single marginal requires solving the entire system, then summing probabilities, so the method complexity is far out of proportion with the difficulty of the problem. On the other hand, true analytical solutions are unavailable for any but the simplest problems, such as constitutive transcription. Inspired by methods in biological and financial stochastic analysis (29,30), we seek to write down generating function solutions that can be evaluated using well-established algorithms, such as numerical quadrature and the fast Fourier transform.

Therefore, we seek to solve systems characterized by the reactions in Eq. 1. We define a set of n transcripts $\{\mathcal{T}_i\}$, whose abundances are represented by the count variables $m_1, \dots, m_n := \mathbf{m}$. The “parent” transcripts are produced with bursts of size B_i ; by intron splicing, they are converted to downstream products and eventually degraded. The splicing reactions can be naturally represented by a directed graph, a discrete structure that encodes causal relationships between states, defined as transcripts in our case. At least one of c_{ij} and c_{ji} must be zero, encoding the intuition that splicing is irreversible and a shorter transcript cannot turn into a longer one. This constraint implies that it is impossible to leave a transcript \mathcal{T}_i and return to it by the process of splicing: the reactions encode an acyclic digraph or *directed acyclic graph* (DAG) (31,32).



METHODS

In this section, we take a modular approach to the CME, and describe how to construct tractable solutions to systems defined by Eq. 1, with the DAG constraint. As outlined in the discussion of scope, we are interested in broad classes of models, and focus on strategies for developing highly general solutions.

The section is organized as follows. For each class of phenomena, we introduce and solve its most general *modular* formulation, but limit the other system components to their simplest form: for general splicing graphs, we discuss geometric bursts; for more complex burst models, we

discuss only a single RNA species. Finally, we describe specific examples, and use the results derived from the general solution to provide qualitative results and constraints.

Definition of the CME

To solve the systems described by Eq. 1, we need to encode them in a CME, which describes probability flux between microstates. We suppose the system has n species, with counts $m_1, \dots, m_n := \mathbf{m}$. The random variables describing the distributions over counts are $\{X_i\}$, with joint probability mass function (PMF) $P(\mathbf{m}, t)$. The reactions in Eq. 1 naively correspond to the following fluxes for each species i :

$$\begin{aligned} R_{tx,i} &= k_i \left[\sum_{z=0}^{m_i} p_{i,z} P(m_i - z, t) - P(m_i, t) \right] \\ R_{deg,i} &= c_{i0} [(m_i + 1)P(m_i + 1, t) - m_i P(m_i, t)] \\ R_{splic,ij} &= c_{ij} [(m_i + 1)P(m_i + 1, m_j - 1, t) - m_i P(m_i, m_j, t)]. \end{aligned} \quad (2)$$

For convenience of notation, Eq. 2 omits the indices of species not involved in a particular reaction. The derivation is provided in section S1.1. The full master equation takes the following generic form:

$$\frac{dP(\mathbf{m}, t)}{dt} = R_{tx} + \sum_{i=1}^n \left[R_{deg,i} + \sum_{j=1}^n R_{splic,ij} \right], \quad (3)$$

i.e., the net flux into any state is the sum of contributions from all reactions for all species. If the burst processes are mutually independent, they can simply be added to yield $R_{tx} = \sum_i R_{tx,i}$, but this is not true in general. Unfortunately, in all but the simplest cases, the PMF is unwieldy to manipulate. Therefore, we convert the CME to the corresponding probability generating function (PGF) $G(x_1, \dots, x_n, t) := G(\mathbf{x}, t)$:

$$G(\mathbf{x}, t) := \sum_{m_1, \dots, m_n} x_1^{m_1} \dots x_n^{m_n} P(\mathbf{m}, t). \quad (4)$$

If we suppose the burst processes *are* independent, this partial differential equation (PDE) takes the following form:

$$\begin{aligned} \frac{\partial G(\mathbf{x}, t)}{\partial t} &= \sum_{i=1}^n k_i (F_i(x_i) - 1)G + \sum_{i=1}^n c_{i0} (1 - x_i) \frac{\partial G}{\partial x_i} \\ &+ \sum_{i,j=1}^n c_{ij} (x_j - x_i) \frac{\partial G}{\partial x_i}, \end{aligned} \quad (5)$$

where F_i is the PGF of the burst random variable B_i . The derivation of this PDE is provided in section S1.2. It is generally easier to treat the logarithm of G ; applying the transformations $u_i := x_i - 1$ and $\varphi := \ln G$ yields the equation:

$$\begin{aligned} \frac{\partial \varphi(\mathbf{u}, t)}{\partial t} &= \sum_{i=1}^n k_i (M_i(u_i) - 1) - \sum_{i=1}^n u_i c_{i0} \frac{\partial \varphi}{\partial u_i} \\ &+ \sum_{i,j=1}^n (u_j - u_i) c_{ij} \frac{\partial \varphi}{\partial u_i}, \end{aligned} \quad (6)$$

where $M_i(u_i) = F_i(1 + u_i)$ is the appropriate transformation of the burst PGF. To compute $P(\mathbf{m}, t)$, we can evaluate $G(\mathbf{x}, t)$ around the n -dimensional unit sphere and take an inverse fast Fourier transform (30,33).

Certain CMEs are isomorphic to SDEs

Before treating the discrete system defined by Eq. 2, we need to establish a crucial connection to a set of stochastic differential equations (SDEs), allowing us to use SDE tools to solve CMEs. Specifically, it is possible to treat X_i as a random variable with the distribution $Poisson(\Lambda_i)$, where Λ_i is the *intensity* of the discrete process. In the general multivariate case, the following identity can be used to relate a set of stochastic processes $\Lambda_1, \dots, \Lambda_n$ with joint distribution $F_{\Lambda_1, \dots, \Lambda_n}$ to the solution of the CME (5,17,34,35):

$$P(\mathbf{m}, t) = \int \prod_{i=1}^n \frac{e^{-\Lambda_i} \Lambda_i^{m_i}}{m_i!} dF_{\Lambda_1, \dots, \Lambda_n}. \quad (7)$$

The *Poisson representation* was introduced by Gardiner (5,35), and successfully applied to modeling regulated gene loci (9,36,37). This representation is always viable, but does not guarantee that the law of Λ_i is a probability: for example, if the variance of marginal random variable X_i is lower than its mean, some probability densities must be negative. Nevertheless, in some cases, we *can* define a relationship between the CME and the “proper” probability laws of SDEs. If the distribution of burst sizes is a Poisson mixture, we can write it down as an equivalent Lévy jump process; for example, the geometric burst size distribution is an exponential-Poisson mixture, and may be interpreted as the Poisson mixture of the underlying compound Poisson process with exponential jumps. This class of models has previously been invoked to explain high trajectory variation observed in living cells (38–40). Thus, if species i undergoes jumps governed by the process $L_{i,t}$, the three prototypical reactions in Eq. 1 take the following form in the continuous worldview:

$$d\Lambda_i = dL_{i,t} - \Lambda_i \sum_{j=0}^n c_{ij} dt + \sum_{j=1}^n c_{ji} \Lambda_j dt, \quad (8)$$

where the left-hand term denotes the evolution of the process, the first term on the right-hand side represents bursting, the second term represents efflux by isomerization and degradation, and the third represents influx by isomerization. If $dL_{i,t} = k_i dt$, this formulation recovers the deterministic reaction rate equations and the well-known Poisson form for constitutive production (34).

This representation has three features that are occasionally useful to solving CME systems. The first is theoretical: the standard properties of Poisson mixtures mean that the moment-generating function (MGF) of the SDE is simply $G(\mathbf{x} - 1, t)$, allowing easy conversion between the two (41,42). Therefore, CME solutions can be used as SDE solutions and vice versa; although we discuss the CME, the solutions generalize to classes of continuous-valued stochastic models explored in the biological (38) and financial (43) literature.

The second feature is qualitative: if the solution to an SDE is available, we can draw conclusions about the isomorphic CME without actually having to solve this CME. For example, the CME in Eq. 9 is isomorphic to the SDE in Eq. 10, as they share the generating function in Eq. 11 (17,44):

$$\begin{aligned} \frac{dP(m, t)}{dt} &= k_1 \sum_{z=0}^m [p_{z,1} P(m - z, t) - P(m, t)] \\ &- \gamma [(m + 1)P(m + 1, t) - mP(m, t)] \end{aligned} \quad (9)$$

$$d\Lambda = dL_t - \gamma \Lambda dt \quad (10)$$

$$G(u, t) = \left(\frac{1 - bu e^{-\gamma t}}{1 - bu} \right)^{\frac{k_1}{\gamma}}, \quad (11)$$

where we suppose the burst distribution is geometric with mean b , the jump distribution is exponential with mean b , bursts arrive according to a Poisson process with rate k_1 , the degradation rate is γ , and the process begins at $\Lambda = X = 0$. Λ_1 is the gamma Ornstein-Uhlenbeck process (29,43–45). Inspired by a highly general result for constitutive transcription, which states that Poisson distributions always remain Poisson for a birth-death process (34), we may reasonably ask whether equivalent results are available for bursty processes. This intuition turns out to be incorrect, and straightforward to disconfirm using SDE results: $G(u, t)$ is not the gamma MGF, so the distribution of the process is not gamma for any finite $t \in (0, \infty)$ —although it does approach a gamma law exponentially fast (44). Therefore, the corresponding Poisson mixture is *not* simply a time-varying negative binomial, and adopting such a model to describe gene expression over time (46) is theoretically problematic.

The third feature is quantitative: the Poisson representation facilitates analysis, both in the notation and the computation. The variance $\mathbb{V}[X_i]$ of the CME is simply given by $\mathbb{V}[\Lambda_i] + \mathbb{E}[\Lambda_i]$, allowing for a more compact representation. Finally, since the generating functions do not assume that the entire system is discrete or continuous, we can use the classes of solutions explored throughout this report to solve *hybrid* systems, with the transcription rate of a CME governed by an SDE. For example, if the transcription rate is given by the gamma Ornstein-Uhlenbeck process, we can solve the hybrid SDE-CME system by appending a “virtual” discrete species that undergoes bursty transcription and splicing (47).

General splicing graphs have analytical solutions

In this subsection, we treat the solutions of CMEs induced by general splicing DAGs. For simplicity, we enforce Eq. 6 holds: all burst processes are independent.

We need to solve Eq. 6, a PDE in n complex variables and one time variable. This is impossible in general. However, using the method of characteristics (30,48,49), we can reduce the PDE’s complex degrees of freedom to n particularly simple coupled ordinary differential equations, which can be represented by a matrix equation:

$$\frac{d\mathbf{U}}{ds} = C\mathbf{U} \text{ s.t. } \mathbf{U}(s=0) = \mathbf{u}, \quad (12)$$

where C is a matrix containing $C_{ij} = c_{ij}$ for all $i \neq j$ and $C_{ii} = -\sum_{j=1}^n c_{ij} - c_{i0}$. We can compute the characteristics by spectral decomposition:

$$U_i(\mathbf{u}, s) = \sum_{j=1}^n e^{-r_j s} (V \cdot \text{Diag}(V^{-1}\mathbf{u}))_{ij} := \sum_{j=1}^n e^{-r_j s} A_{ij}(\mathbf{u}), \quad (13)$$

where r_j are the (strictly nonnegative) eigenvalues of $-C$ and each column of matrix V contains an eigenvector of C . The derivation of this solution procedure is provided in section S2. Once the functional form of U_i is available, it is straightforward to compute the solution of Eq. 6 by quadrature:

$$\varphi(\mathbf{u}, t) = \int_0^t \sum_{i=1}^n k_i M_i(U_i(\mathbf{u}, s) - 1) ds. \quad (14)$$

Moments of the splicing graph solutions are tractable by matrix operations

To discuss specific results, such as the correlations between transcripts, we assume that exactly one parent transcript exists, and all other transcripts are produced from it by splicing. Consistently with the conventional model (25,30,50) we assume it is produced in geometrically distributed bursts.

This corresponds to defining \mathcal{T}_1 as the parent transcript and setting $M_1(u_1) = M(u_1) = \frac{1}{1-bu_1}$ in Eq. 6, with all $k_j, j \neq 1$ defined as zero.

To analyze this system, we rewrite Eq. 13 in terms of its marginal components:

$$U_1(\mathbf{u}, t) = \sum_{i=1}^n u_i \psi_i(t) \text{ s.t. } \psi_i = \sum_{k=1}^n a_{i,k} e^{-r_k s}. \quad (15)$$

This form is most convenient for analyzing the summary statistics of a few marginals at a time. If the spectral decomposition is available, converting to this representation amounts to evaluating the coefficient matrix A once: $a_{i,k} = A_{1k}(\delta_{ij}) = (V \cdot \text{Diag}(V^{-1}\delta_{ij}))_{1j}$, where δ_{ij} is the Kronecker delta vector with a 1 in position i and 0 elsewhere.

The moments of the marginal distribution of species i can be computed directly from the derivatives of the marginal MGF of the formal *continuous* system with all complex arguments in \mathbf{u} set to zero. We begin by considering U_1 as the sum in Eq. 15. To marginalize with respect to all $j \neq i$, we simply set all u_j to zero, obtaining $U_1 = u_i \psi_i(t)$. To compute the lower moments of a single species, we take the derivative with respect to u_i and evaluate it at $u_i = 0$ with the help of Eq. 15:

$$\begin{aligned} \mathbb{E}[\Lambda_i] &= e^\varphi \frac{\partial}{\partial u_i} k_1 \int_0^\infty \left[\frac{1}{1 - bu_i \psi_i(s)} - 1 \right] ds \Bigg|_{u_i=0} \\ &= k_1 b \int_0^\infty \psi_i(s) ds = k_1 b \sum_{k=1}^n \frac{a_{i,k}}{r_k}. \end{aligned} \quad (16)$$

Per the standard properties of mixed Poisson distributions (41), the value of μ_i is identical for the underlying continuous process and the derived discrete process.

The covariances can be computed directly from the derivatives of the marginal PGF with $u_q = 0$ for all $q \neq i, l$. By construction, $U_1(u_i, u_l; s) = u_i \psi_i(s) + u_l \psi_l(s)$, where each ψ is the exponential sum corresponding to the marginal of the species in question. Taking the partial derivatives yields the following equation:

$$\text{Cov}(\Lambda_i, \Lambda_l) = 2k_1 b^2 \sum_{j,k=1}^n \frac{a_{i,j} a_{l,k}}{r_j + r_k}. \quad (17)$$

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 (41). The marginal variances can be found by plugging in $l = i$, and the standard properties of Poisson mixtures (41) allow conversion to the discrete domain, with $\mathbb{V}[X_i] = \mathbb{V}[\Lambda_i] + \mu_i$. 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. These lower moments are straightforward to compute, but do not appear to have an easily amenable analytical form for general graphs.

General burst distributions are tractable and encode physics with no free parameters

The relevance of CME to modern transcriptomic experimental data is tempered by the simplicity of tractable models. The model we have presented so far can describe the splicing cascade of a single gene, but does not naturally extend to multigene networks. Yet we know that genes often belong to coexpression modules that are identifiable by similarity metrics (2,21,23). 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 (51), 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 (52). 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.

For convenience of notation, we omit any discussion of the details of the downstream splicing networks, as the results derived above hold with no change. We have thus far assumed the $R_{i,c,i}$ in Eq. 2 are *independent*, i.e., their bursts are unsynchronized. However, more sophisticated models can be built, and encode gene-gene correlations. Most generally, the differential of the log-PGF in Eq. 6 takes the following functional form:

$$\frac{\partial \varphi(\mathbf{u}, t)}{\partial t} = \sum_{\ell=1}^n \sum_{q=1}^{\omega} k_{\ell,q} (M_{\ell,q}(\mathbf{U}(\mathbf{u}, t)) - 1). \quad (18)$$

The full derivation is provided in section S1.4, and relies on iterative application of Cauchy products to conditional PGFs. This rather formal expression states that up to n species in the system may be cotranscribed in a single module. For a particular module of ℓ genes with synchronized transcription times, there are $\omega = \binom{n}{\ell}$ possible combinations of species that can be cotranscribed, indexed by q . The independent case emerges from setting all $k_{\ell,q}$ to zero for $\ell > 1$. The vector defining $U_1, \dots, U_n := \mathbf{U}$ is modular, as it is independent of bursting dynamics.

Example: Two-promoter bursty model, no synchronization

To begin, consider a system with two independent promoters that fire reactions $\emptyset \xrightarrow{k_{1,1}} B_1 \times \mathcal{T}_1$ and $\emptyset \xrightarrow{k_{1,2}} B_2 \times \mathcal{T}_2$. If \mathcal{T}_1 can be converted to \mathcal{T}_2 , this model can describe an internal promoter (53) that generates molecules that cover only a part of the gene. In this formulation, the following special case of Eq. 6 holds:

$$\frac{\partial \varphi(\mathbf{u}, t)}{\partial t} = k_{1,1} (M_1(U_1(\mathbf{u}, t)) - 1) + k_{1,2} (M_2(U_2(\mathbf{u}, t)) - 1). \quad (19)$$

In the parlance of Eq. 18, we set $n = 2$ and $k_{2,1} = 0$. Therefore, the stationary log-PGF takes the simple form:

$$\varphi(\mathbf{u}, t) = \int_0^t [k_{1,1} (M_1(U_1(\mathbf{u}, s)) - 1) + k_{1,2} (M_2(U_2(\mathbf{u}, s)) - 1)] ds, \quad (20)$$

i.e., the joint distribution of RNA can be obtained with a single application of quadrature.

If U_1 and U_2 are disjoint (e.g., \mathcal{T}_1 and \mathcal{T}_2 are products from two different genes), this expression reduces to the trivial case $\varphi(\mathbf{u}) = \varphi_1(u_1) + \varphi_2(u_2)$: intuitively, independent transcription processes produce statistically independent distributions. If only a single parent transcript exists ($U_1 = U_2 = U$), the model can represent a particular class of multistate promoters with two distinct short-lived active states, recapitulating the unsynchronized model in section S1.3. Finally, if the burst distributions are identical ($M_1 = M_2 = M$), the system becomes equivalent to a one-locus system with burst frequency $k_{1,1} + k_{1,2}$. This accords with the superposition property of the Poisson process driving transcription.

Example: Two-gene bursty model, with burst time synchronization

We continue by considering the instructive model of two genes influenced by the same regulator: the active periods of these genes are synchronized in time. However, the burst sizes are not coupled, and may indeed come from different distributions. Such a process has the transcription reaction $\emptyset \xrightarrow{k_{2,1}} B_1 \times \mathcal{T}_1 + B_2 \times \mathcal{T}_2$, where each \mathcal{T}_i may be degraded or processed into further downstream species with total efflux rate r_i . In the parlance of Eq. 18, we set $k_{\ell,q}$ to 0 for $\ell = 1$.

Assuming that $B_1 \sim \text{Geom}(b_1)$ and $B_2 \sim \text{Geom}(b_2)$ —i.e., the marginals are consistent with the standard bursty model—we can exploit the independence of B_1 and B_2 to write down the joint distribution:

$$M(\mathbf{U}(\mathbf{u}, t)) = M_1(U_1(\mathbf{u}, t))M_2(U_2(\mathbf{u}, t)), \quad (21)$$

which follows immediately from standard MGF identities. We propose a physical model leading to this burst size distribution in section S1.3. The expression in Eq. 18, with Eq. 21 used as the transcriptional burst size distribution, has the following solution:

$$\begin{aligned} \varphi &= k_{2,1} \int_0^t [M_1(U_1(\mathbf{u}, s))M_2(U_2(\mathbf{u}, s)) - 1] ds \\ &= k_{2,1} \int_0^t \left[\frac{1}{(1 - b_1 U_1(\mathbf{u}, s))(1 - b_2 U_2(\mathbf{u}, s))} - 1 \right] ds. \end{aligned} \quad (22)$$

This expression is numerically integrable, but does not afford an analytical solution. We can obtain its lower moments by differentiating the log-PGF. Defining $U_i = u_i \psi_i$ and $U_l = u_l \psi_l$ for two transcript species generated from distinct genes, we yield:

$$\begin{aligned} \text{Cov}(\Lambda_i, \Lambda_l) &= k_{2,1} b_1 b_2 \int_0^\infty \psi_i \psi_l ds \\ &= k_{2,1} b_1 b_2 \sum_{j,k=1}^n \frac{a_{i,j} a_{l,k}}{r_j + r_k}. \end{aligned} \quad (23)$$

Usefully, this expression is agnostic of the actual identity of i, l , so the expression holds for *any* of the species downstream of \mathcal{T}_1 and \mathcal{T}_2 . Finally, if we would like to compute the covariance between the two parent species, we simply plug in $\psi_i = e^{-r_1 s}$ and $\psi_l = e^{-r_2 s}$:

$$\text{Cov}(\Lambda_1, \Lambda_2) = \frac{k_{2,1} b_1 b_2}{r_1 + r_2}. \quad (24)$$

This covariance corresponds to the following Pearson correlation coefficient:

$$\begin{aligned} \rho &= \frac{\text{Cov}(\Lambda_1, \Lambda_2)}{\sigma_1 \sigma_2} = \frac{\sqrt{r_1/r_2}}{1 + r_1/r_2} \\ &\times \sqrt{\frac{1}{(1 + 1/b_1)(1 + 1/b_2)}}. \end{aligned} \quad (25)$$

The first term achieves a global maximum of $1/2$ at $r_1 = r_2$. 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.

Example: Two-gene bursty model, with perfect burst time and size synchronization

Conversely, we can consider the two-gene problem assuming that the burst distributions are identical and perfectly 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:

$$\begin{aligned} F(x_1, x_2) &= \sum_{i,j} x_1^i x_2^j p_{i,j} = \sum_{i,j} x_1^i x_2^j p_i \delta(i-j) \\ &= \frac{1}{1 + b - bx_1 x_2}. \end{aligned} \quad (26)$$

Upon inserting the characteristics and taking the requisite partial derivatives, we yield the following correlation structure:

$$\rho = \frac{\sqrt{r_1/r_2}}{1 + r_1/r_2} \times \left(1 + \frac{b}{b+1}\right). \quad (27)$$

As in the case of uncoupled gene sizes reported in Eq. 25, the first term is at most 1/2. The second term asymptotically approaches 2 as $b \rightarrow \infty$. 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.

Example: Two-gene bursty model, with partial burst time and size synchronization

The models described above are useful for understanding limiting cases, but somewhat restrictive: the model solved in Eq. 19 enforces $\rho = 0$, the synchronized burst time model in Eq. 22 enforces $\rho \in (0, 1/2)$, and the perfectly synchronized burst size model in Eq. 26 requires burst sizes to be identical. Thus, we need to construct models consistent with observations, recapitulating both the multigene correlation structure and the diversity of burst sizes evident from marginals.

This is most easily tractable using the continuous formulation: we can impose perfect correlation between Lévy jump sizes. Conceptually, we suppose the overall driving process L_t for N genes is given by a compound Poisson process with average jump size b and jump distribution $J \sim \text{Exp}(b^{-1})$. The driving process for a single gene i is given by $L_{i,t} = w_i L_t$, with average jump size bw_i and $\sum_{i=1}^N w_i = 1$. We propose a physical model leading to this burst size distribution in section S1.3. Since all jump sizes are scalar multiples of each other, they are perfectly correlated. The joint jump distribution has the following MGF:

$$\begin{aligned} M(\mathbf{u}) &= \mathbb{E} \left[\exp \left(\sum_{i=1}^N u_i J_i \right) \right] = \mathbb{E} \left[\exp \left(J \sum_{i=1}^N u_i w_i \right) \right] \\ &= \frac{1}{1 - b \sum_{i=1}^N u_i w_i}, \end{aligned} \quad (28)$$

where the third equality stems from recognizing that the joint MGF is simply the univariate exponential MGF evaluated at $\sum_{i=1}^N u_i w_i$. If we only consider the parent transcripts with efflux rates r_i and characteristics $U_i = u_i e^{-r_i t}$, we yield the following stationary log-PGF:

$$\varphi(\mathbf{u}) = k_{N,1} \int_0^\infty \left[\frac{1}{1 - b \sum_{i=1}^N u_i w_i e^{-r_i s}} - 1 \right] ds. \quad (29)$$

Finally, setting $N = 2$, we can take derivatives of the PGF and compute the correlation structure:

$$\text{Cov}(\Lambda_1, \Lambda_2) = \frac{2k_{2,1} b^2 w_1 w_2}{r_1 + r_2} \quad (30)$$

$$\rho = \frac{2\sqrt{r_1/r_2}}{1 + r_1/r_2} \times \sqrt{\frac{1}{\left(1 + \frac{1}{bw_1}\right) \left(1 + \frac{1}{bw_2}\right)}} \in (0, 1). \quad (31)$$

As above, this model is appealing since it has no free parameters: the gene-gene correlations are not imposed by an ad hoc correlation parameter, but emerge from the model structure itself. Therefore, we can evaluate the model's performance by comparing true intergene correlations to ones computed based only on the marginals.

Interestingly, single-gene systems with rapid splicing recapitulate the multigene functional form. Consider a source species \mathcal{T}_0 , which is produced in geometrically distributed bursts and converted to species $\{\mathcal{T}_i\}$, with $i = 1, \dots, N$, at rates β_i . These transcripts are degraded at rates r_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 by a Lévy process with jumps of average size $b\beta_i/\beta$, where b is the underlying \mathcal{T}_0 burst size, $\beta := \sum_i \beta_i$, and the ratio β_i/β is $\mathcal{O}(1)$. We define corresponding weights $w_i := \beta_i/\beta$; by definition, $\sum_i w_i = 1$. This yields:

$$U_0(\mathbf{u}, t) = C e^{-\beta t} + \sum_{i=1}^N \frac{\beta_i u_i}{\beta - r_i} e^{-r_i t} \rightarrow \sum_{i=1}^N u_i w_i e^{-r_i t}, \quad (32)$$

where C is a constant in the solution of the characteristic $U_0(\mathbf{u}, t)$ whose value becomes immaterial as β increases. Plugging in the characteristic:

$$M(U_0(\mathbf{u}, t)) = \frac{1}{1 - b U_0(\mathbf{u}, t)} = \frac{1}{1 - b \sum_{i=1}^N u_i w_i e^{-r_i t}}, \quad (33)$$

which recapitulates the form of the integrand in Eq. 29. Therefore, the multigene model can be used to describe transcripts arising from a single rapidly processed, unobserved parent molecule, yielding positive, but otherwise unconstrained, correlation between the downstream species.

Example: Anti-correlated two-gene bursty model, with burst time synchronization

Thus far, we have considered the problem of describing synchronized genes, and proposed three models that can produce positive correlations. Putting aside the problem of positing a specific physical mechanism, we can ask whether *any* joint burst distribution can produce negative correlations in molecule counts, *despite* perfect synchronization of burst events. Considering the cross moment of mRNA produced at two synchronized loci, with generic joint burst generating function M , we find:

$$\mathbb{E}[\Lambda_1 \Lambda_2] = k_{2,1} \int_0^\infty \frac{d^2 M}{du_1 du_2} ds + \mu_1 \mu_2, \quad (34)$$

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

$$\frac{d^2 M}{du_1 du_2} = \frac{d^2 M}{dU_1 dU_2} \frac{dU_1}{du_1} \frac{dU_2}{du_2} = \mathbb{E}[B_1 B_2] e^{-(r_1+r_2)s}, \quad (35)$$

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

$$\text{Cov}(\Lambda_1, \Lambda_2) = k_{2,1} \frac{\mathbb{E}[B_1 B_2]}{r_1 + r_2} = \frac{k_{2,1}}{r_1 + r_2} (\rho \sigma_{B_1} \sigma_{B_2} + \mu_{B_1} \mu_{B_2}), \quad (36)$$

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

$$\text{Cov}(\Lambda_1, \Lambda_2) \geq \frac{k_{2,1}}{r_1 + r_2} (\mu_{B_1} \mu_{B_2} - \sigma_{B_1} \sigma_{B_2}). \quad (37)$$

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 $\sigma_{B_1}^2 = \sigma_{B_2}^2 = \alpha\theta^2$. Therefore, the covariance of the Poisson intensities takes the following form:

$$\begin{aligned} \text{Cov}(\Lambda_1, \Lambda_2) &= \frac{k_{2,1}}{r_1 + r_2} (\rho\alpha\theta^2 + \alpha^2\theta^2) \\ &= \frac{k_{2,1}\theta^2}{r_1 + r_2} (\rho\alpha + \alpha^2), \end{aligned} \quad (38)$$

which achieves $\text{Cov}(\Lambda_1, \Lambda_2) < 0$ 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 $\rho \in (-1, 1)$ (54,55). Applying the Cauchy-Schwartz inequality to the marginals guarantees that the joint MGF of the correlated bivariate gamma distribution exists (56). This demonstrates the existence of continuous moving average processes with negative stationary correlation, driven by a common 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.

Certain models of heterogeneity are analytically tractable

Thus far, we have reported a toolbox of models that can represent Markovian systems with fairly general splicing graphs and burst distributions, and demonstrated that their solutions exist and are computable. By design, these models by themselves describe homogeneous populations of cells, with deterministic parameters. Using standard statistical approaches, they can be reassembled to describe heterogeneity in parameters (57):

$$P(\mathbf{m}, t) = \int_{\Theta} P(\mathbf{m}, t; \theta) df_{\theta} = \mathbb{E}_{\Theta}[P(\mathbf{m}, t; \theta)] \quad (39)$$

$$G(\mathbf{x}, t) = \int_{\Theta} G(\mathbf{x}, t; \theta) df_{\theta} = \mathbb{E}_{\Theta}[G(\mathbf{x}, t; \theta)]. \quad (40)$$

Equation 39 is essentially the definition of a mixture model: the parameter vector θ may itself be distributed according to the statistical law f_{θ} over a state space Θ . To determine the resulting probability distribution, we need to integrate with respect to this law. By linearity, we can exchange the generating function sum in Eq. 4 and the integral in Eq. 39 to obtain Eq. 40.

This expression is extremely generic; the base case of a homogeneous cell population with parameter vector $\bar{\theta}$ is obtained by setting $df_{\theta} = \delta(\theta - \bar{\theta})$, where δ is a multivariate Dirac delta functional. Unfortunately, analytical solutions are only available in several highly restrictive cases.

Example: A finite number of cell types

Multimodality in single cells can emerge from the existence of multiple long-lived subpopulations, which are conventionally fit to a finite mixture distribution (58). We can formalize this model for an arbitrary number of populations. Consider a population of cells with J disjoint “cell types,” indexed by j . Each cell type has the parameter vector θ_j and relative abundance w_j . The probability distribution of the entire population can be computed by the law of total probability, conditioning on the cell type, or by setting $df_{\theta} = \sum_{j=1}^J w_j \delta(\theta - \theta_j)$ in Eqs. 39 and 40:

$$P(\mathbf{m}, t) = \sum_{j=1}^J w_j P(\mathbf{m}, t; \theta_j) G(\mathbf{x}, t) = \sum_{j=1}^J w_j G(\mathbf{x}, t; \theta_j). \quad (41)$$

This formulation can produce J -modal distributions. This reduces to the standard model when $J = 1$ or when all θ_j are identical.

Correlations can emerge from the existence of multiple cell types. For simplicity, we set $J = 2$ and $n = 2$, and suppose the molecule distributions are independent Poisson, whether as an effect of constitutive production or as the limit of a bursty distribution. For gene i and cell type j , the average expression is $\lambda_{i,j}$, implying that the stationary distribution takes the following form:

$$\begin{aligned} G(\mathbf{u}) &= \sum_{j=1}^2 w_j G(\mathbf{u}; \lambda_{1,j}, \lambda_{2,j}) = \sum_{j=1}^2 w_j G(u_1, \lambda_{1,j}) G(u_2, \lambda_{2,j}) \\ &= \sum_{j=1}^2 w_j e^{\lambda_{1,j} u_1} e^{\lambda_{2,j} u_2}. \end{aligned} \quad (42)$$

The correlation can be found by differentiation. Setting $w_1 = w_2 = \frac{1}{2}$ and supposing $\lambda_{i,1} = 0$ (i.e., cell type 1 does not express either gene), we yield:

$$\rho = \sqrt{\frac{\lambda_{1,2} \lambda_{2,2}}{(\lambda_{1,2} + 2)(\lambda_{2,2} + 2)}} \in (0, 1). \quad (43)$$

Supposing now $\lambda_{1,2} = \lambda_{2,1} = 0$ (i.e., gene 1 is a perfectly specific marker for cell type 1 and gene 2 is a marker for cell type 2):

$$\rho = -\sqrt{\frac{\lambda_{1,1} \lambda_{2,2}}{(\lambda_{1,1} + 2)(\lambda_{2,2} + 2)}} \in (-1, 0). \quad (44)$$

Therefore, correlations can emerge even in the absence of transcriptional synchronization. To account for these correlations, it is necessary to build mixture models, investigate living systems with low cell type diversity, or use ad hoc filtering to extract putative homogeneous cell populations.

Example: Extrinsic noise in burst frequency

Even within a homogeneous cell type, the parameter values may not be perfectly synchronized. We can thus define *extrinsic* noise, which encodes

the stochastic parameter distribution (24,59). This description presupposes that the noise, if generated by a stochastic process, evolves much slower than the transcriptional dynamics (47). There does not appear to be a general solution for parameter mixing, but a set of solutions is available for systems with random burst frequencies. Suppose there is a single parent transcript with a burst frequency described by the random variable K_i . We can immediately write down a solution for the overall count PGF:

$$\begin{aligned} G(\mathbf{x}, t) &= \int_0^\infty G(\mathbf{x}, t; k_i) df_{K_i} = \int_0^\infty e^{k_i \varphi^*(\mathbf{x}, t)} df_{K_i} \\ &= M_{K_i}[\varphi^*(\mathbf{x}, t)], \end{aligned} \quad (45)$$

where M_{K_i} is the MGF of the burst frequency distribution and φ^* is a nominal log-PGF with unity burst frequency. The degenerate case is recovered by constraining K_i to a Dirac delta distribution at k_i , with MGF $M_{K_i}(z) = e^{k_i z}$. Equation 45 immediately extends to integrals of Eq. 18 by defining a multivariate burst frequency distribution and computing its MGF.

Example: Extrinsic noise in burst size

Burst size modulation has been heavily implicated in the regulation of cell size (60–62). For systems with a distribution of *static* cell sizes, not governed by the cell cycle, we can at least formally postulate a model:

$$G(\mathbf{x}, t) = \int_0^\infty G(\mathbf{x}, t; \mathbf{b}(z)) df_Z(z), \quad (46)$$

where \mathbf{b} is the vector of average burst sizes, governed by a common *univariate* random variable Z with realizations z , e.g., $b_i = zc_i$. Z thus models the “cell size” in a mechanistic fashion. Unfortunately, despite this relevance to physiology, this general form is intractable: the burst size parameter cannot be “factorized” out. However, certain special cases can be connected to previous work.

Consider the case of a single mRNA species produced in geometric bursts with a random cell-specific burst size. If the inverse burst size distribution is given by the law \mathcal{Z} , the distribution of mRNA is given by the \mathcal{Z} -gamma-Poisson PMF. Only a few \mathcal{Z} -gamma distributions have been studied in depth. If \mathcal{Z} is gamma (i.e., the burst size is inverse-gamma distributed), gamma-gamma compounding yields a beta distribution of the second kind (63), with a rather complicated MGF available in terms of Lauricella functions (64).

Transient burst dynamics are solvable

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 with the mRNA lifetimes (1,65). Usually, the regulatory events underpinning these processes are modeled by variation in DNA-localized transcriptional parameters (66–69).

By examining Eq. 18, it is easy to see that the current framework can be extended to *any* deterministic variation in k and M , with solutions available at arbitrary times t :

$$\begin{aligned} \varphi(\mathbf{u}, t) &= \int_0^t \sum_{\ell=1}^n \sum_{q=1}^{\binom{n}{\ell}} k_{\ell,q}(s) (M_{\ell,q}(\mathbf{U}(\mathbf{u}, s), s) - 1) ds. \end{aligned} \quad (47)$$

Therefore, burst frequencies, burst distributions, and even the coexpression modules can vary, continuously or discontinuously, as long as the variation is deterministic. This can be exploited to model a variety of phenomena, such as variation in gene copy numbers over the cell cycle

(62,66) and the concentration of high-abundance regulators not coupled to the mRNA under investigation. If the reaction rates within \mathbf{U} change over time, the generating function PDE becomes intractable; however, some simple models, such as piecewise constant, can be solved by splitting the integral.

We have further assumed $m_i(t=0) = 0$ for all species i . However, if nonzero molecule counts are present at $t = 0$, it is straightforward to compute the resulting log-PGF by separately defining the homogeneous generating function φ^h with $m_i(t=0) = 0$ and the generating function of the initial condition φ^{init} :

$$\varphi(\mathbf{u}, t) = \varphi^h(\mathbf{u}, t) + \varphi^{init}(\mathbf{U}(\mathbf{u}, t)) \quad (48)$$

$$\varphi(\mathbf{u}, t) = \varphi^h(\mathbf{u}, t) + \sum_{i=1}^n m_i^{(0)} \ln(1 + U_i(\mathbf{u}, t)), \quad (49)$$

where U_i are the characteristics. Eq. 48 relates the general form of the conditional distribution, whereas Eq. 49 produces the particular form with deterministic initial molecule counts $m_i^{(0)}$, as discussed elsewhere (70). 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.

Example: Cell cycling

Using the conditioning relation in Eq. 49, we can solve models of certain cell-cycle phenomena. Suppose that a “cell cycle” consists of two stages with durations Δ_1 and Δ_2 and parameters $b^{(j)}, \gamma^{(j)}, k_1^{(j)}$ in stage j . The PGF of the joint transcript distribution at the end of the first stage, $t = \Delta_1$, is given by Eq. 48, yielding the PGF $G^{(1)}(\mathbf{u})$. Now supposing the stage transition involves the binomial partitioning of all molecules, the distribution of mRNA upon entering the second stage is $G^{(1)}(1 + \mathbf{u}/2)$. This identity results from the law of total expectation, and amounts to asserting that the probability of retaining each molecule is 1/2, with a per-molecule Bernoulli retention PGF of $\frac{1}{2}(1+x)$. This result has been derived in previous models of cell cycling (25,62) and proposed in more generality in our recent discussion of technical noise (16).

Considering the specific example of a single transcript with initial PGF $G^{(0)}$, we can adapt Eq. 11 (44) to yield:

$$G^{(1)}(u, \Delta_1) = G^{(0)} \left(u e^{-\gamma^{(1)} \Delta_1} \right) \left(\frac{1 - b^{(1)} u e^{-\gamma^{(1)} \Delta_1}}{1 - b^{(1)} u} \right)^{\frac{k_1^{(1)}}{\gamma^{(1)}}}. \quad (50)$$

This is the generating function immediately before partitioning. The generating function immediately after simply inserts $\frac{1}{2}(2+u)$ in place of u to account for partitioning. Finally, to compute the generating function at the end of the cycle, we multiply through by the transient PGF and insert the characteristic $u e^{-\gamma^{(2)} t}$ into the initial condition:

$$\begin{aligned} G^{(2)}(u, \Delta_1 + \Delta_2) &= G^{(0)} \left(\frac{1}{2} \left(2 + u e^{-\gamma^{(2)} \Delta_2} \right) e^{-\gamma^{(1)} \Delta_1} \right) \\ &\times \left(\frac{1 - b^{(1)} \frac{1}{2} \left(2 + u e^{-\gamma^{(2)} \Delta_2} \right) e^{-\gamma^{(1)} \Delta_1}}{1 - b^{(1)} \frac{1}{2} \left(2 + u e^{-\gamma^{(2)} \Delta_2} \right)} \right)^{\frac{k_1^{(1)}}{\gamma^{(1)}}} \\ &\times \left(\frac{1 - b^{(2)} u e^{-\gamma^{(2)} \Delta_2}}{1 - b^{(2)} u} \right)^{\frac{k_1^{(2)}}{\gamma^{(2)}}}. \end{aligned} \quad (51)$$

Although solutions of this type are somewhat unwieldy to write down explicitly, they *can* always be computed by composition of functions for the one-species system.

The general solution, with arbitrary burst distributions, splicing networks, and deterministic regulatory dynamics, can be computed by working backward from the last stage of cell cycle and incrementally adding “initial condition” contributions from previous stages. As an illustration, we define a system with zero molecules at $t = 0$, with a bursty promoter that produces a single parent isoform with the characteristic $U_1(\mathbf{u}, s)$. For convenience, we assume the splicing and degradation are time independent. We can immediately write down the generating function for the system state before partitioning at $t = \Delta_1 + \Delta_2$:

$$\varphi^{(2)}(\mathbf{u}, \Delta_1 + \Delta_2) = \left(k_1^{(1)} \int_0^{\Delta_1} \left[\frac{1}{1 - b^{(1)} U_1\left(\frac{1}{2}(2 + \mathbf{U}(\mathbf{u}, \Delta_2)), s\right)} - 1 \right] ds \right) + \left(k_1^{(2)} \int_0^{\Delta_2} \left[\frac{1}{1 - b^{(2)} U_1(\mathbf{u}, s)} - 1 \right] ds \right). \quad (52)$$

This generalizes the systems studied previously (62) by relaxing the assumption of timescale separation between transcripts, while maintaining the assumption of bursty gene dynamics; transient coupling between cell size and burst size (such as the exponential form in (62)) can be incorporated by appropriately defining a time-varying burst distribution in Eq. 47. The solution has previously been extended to describe Erlang-distributed stochastic time intervals (25,71). This requires defining a CME coupled to a multistate Markov chain governing the transcriptional parameters and yields a series of coupled PDEs that take a form reminiscent of multistate promoter equations (72), but are not generally tractable by quadrature.

RESULTS

Ultimately, we would like to use these solutions to fit real data, and represent entire data sets using a small set of physically interpretable parameters for each gene, potentially with some higher-level structure encoding cell types (as in Eq. 41) or gene-gene synchronization (as in Eq. 29). Unfortunately, the experimental and computational infrastructure to do this does not exist yet: as we discuss at length in section S5, single-cell, single-molecule, full-length sequencing methods are in their nascence, the structure of splicing processes is not well characterized on a genome-wide scale, and both biology and sequencing involve obscure sources of noise.

However, we *can* take the first steps in this direction by exploiting some of the simpler results to explain correlations in real data. In Eqs. 33 and 29, we report two models that give positive correlations for coexpressed transcripts. We cannot yet use these models to recapitulate entire data sets. The physics are too complex to explicitly describe: the model is a priori misspecified because we disregard other sources of noise. Furthermore, the inference procedure requires some care, as the joint distributions are too large to compute. If we try to fit distributions two at a time, the results will be inconsistent (e.g., fitting genes 1 and 2 will

yield an estimated \hat{b}_1 different from the one obtained by fitting genes 1 and 3), and still suffer from model misspecification.

We can sacrifice some statistical power but improve interpretability by treating the genes one at a time. The models’ correlation structure, given in Eq. 31, has no free parameters: it is fully determined by the marginal distributions. Given marginal estimates of gene-specific parameters (straightforward to compute by maximum likelihood estimation with the negative binomial law), we can *predict*

the correlation structure and compare predictions to experimental ground truth. We interpret these predictions as *upper bounds* in the absence of all other noise: if additional sources of stochasticity are present, the correlations should degrade relative to the model. As the correlations predicted by Eq. 31 are strictly less than 1, they are nontrivial.

To perform this analysis, we obtained data from the recent FLT-seq (full-length transcript sequencing by sampling) protocol (73). As this experimental technique has molecular and cellular barcodes, the data are interpretable as discrete transcript counts sampled from a distribution. To minimize transient effects, such as cell cycling and differentiation, we selected a data set generated from cultured mouse stem cells. To limit biological heterogeneity due to discrete cell subpopulations (as in Eq. 41), we filtered for cell barcodes corresponding to the activated cell subset (136 barcodes) according to the authors’ annotations. In all downstream analyses, we treated this filtered data set as biologically homogeneous up to intrinsic stochasticity.

The FLT-seq protocol produces full-length reads, which can be used to discover new isoforms, but does not reveal causal relationships between those isoforms. Nevertheless, we can use the tools of discrete mathematics to partially infer these relationships. Splicing removes introns, but cannot insert them. We can use this relationship to constrain the splicing DAG: if transcript \mathcal{T}_j can be obtained by removing part of the sequence in transcript \mathcal{T}_i , there must be a path from \mathcal{T}_i to \mathcal{T}_j . On the other hand, if \mathcal{T}_i contains the sequence I_i but omits the sequence I_j , whereas \mathcal{T}_j contains the sequence I_j but omits the sequence I_i , the transcripts are *mutually exclusive* and must be generated from the parent transcript by disjoint processes.

For each gene, we enumerate the transcripts observed in the data and split them into elementary intervals, contiguous stretches that are either present or absent in each transcript (denoted by the colors in Fig. 1 a). These elementary

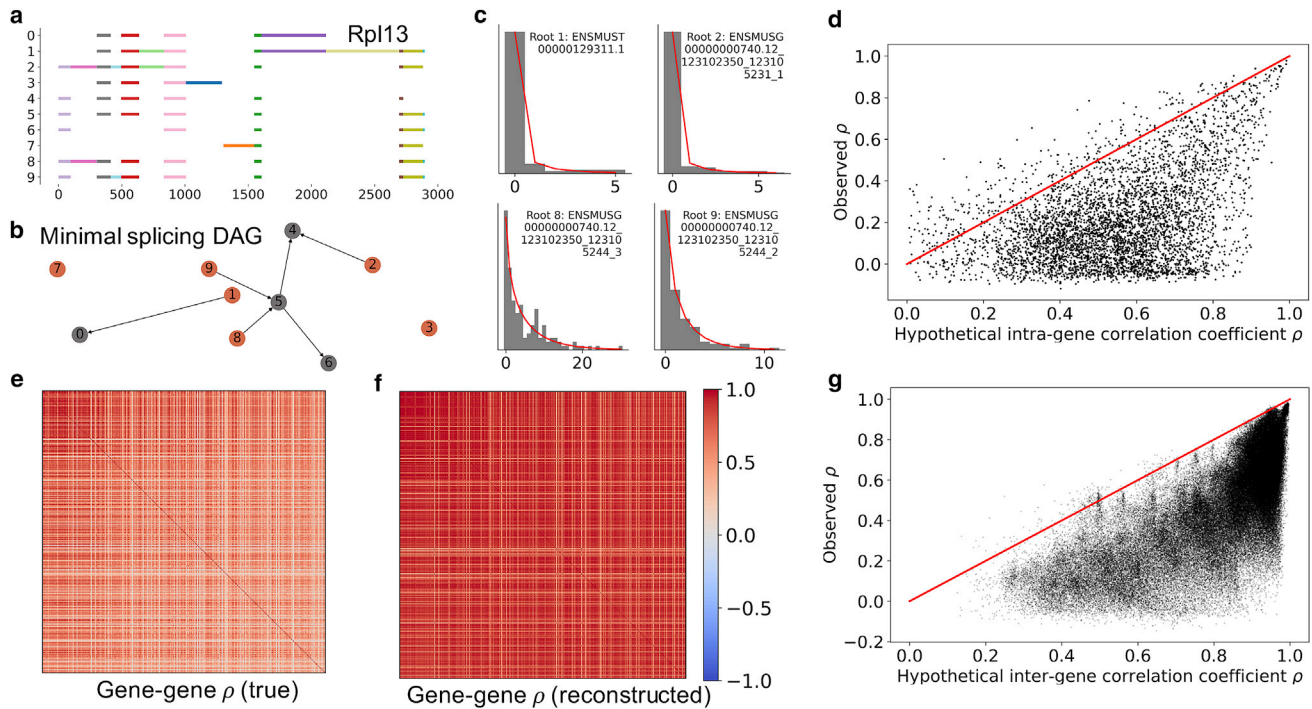


FIGURE 1 Leveraging the synchronized burst model to predict transcript-transcript correlations. (a) By inspecting exon coexpression structures in long-read sequencing data, we can split genes into elementary intervals. (b) Although sequencing data are not sufficient to identify the relationships between various transcripts, they can provide information about “roots” of the splicing graph (highlighted in orange), which must be produced from the parent transcript by mutually exclusive processes. (c) The root transcript copy-number distributions are well described by negative binomial laws (gray histograms, raw marginal count data; red lines, fits). (d) The intrinsic noise model is not sufficient to accurately predict transcript-transcript correlations, but does serve as a nontrivial upper bound: few sample correlations exceeded the model-based predictions obtained from Eq. 31 (points, transcript-transcript correlation matrix entries for mutually exclusive “root” transcripts of a single gene; red line, theory/experiment identity line). (e) The highest-expressed transcripts across the top 500 genes show distinctive, and generally positive correlation patterns. (f) We can use an analogous model to predict and reconstruct the correlation matrix based solely on marginal data. (g) As before, the model (Eq. 31) is not sufficient to accurately predict gene-gene correlations, but provides an effective and nontrivial upper bound (points, transcript-transcript correlation matrix entries; red line, theory/experiment identity line). To see this figure in color, go online.

intervals constrain the relationships between transcripts, and we can use their presence or absence in each transcript to construct an accessibility graph. The internal structure of this graph is underspecified, but immaterial: the negative binomial model implied by the generating functions in Eqs. 33 and 29 describes the roots, mutually exclusive transcripts that must be generated directly from the parent transcript (indicated in orange in Fig. 1 b, and solved in Eq. 33). We fit the distributions of these roots, discarding any data that are underdispersed (variance lower than mean), overly sparse (fewer than five molecules in the entire data set), or fail to converge to a fit. The satisfactory fits for the sample gene Rpl13 are shown in Fig. 1 c.

The negative binomial fit yields transcript-specific burst sizes bw_i and efflux rates r_i (nondimensionalized by setting burst frequency to unity). We plug these quantities into Eq. 31, compute hypothetical correlations ρ_{theo} , and compare them to sample correlations ρ_{samp} in Fig. 1 d. These results represent the 4,885 nontrivial correlation matrix entries between 1,978 transcripts from 500 genes. A total of 302 transcripts were rejected due to underdispersion, 542 due to sparsity, and 100 due to poor fits. The theoretical

constraint (sample correlation equal to or lower than predicted correlation) was met in 4,606 cases (94.3%).

The results suggest that the model is not sufficient to recapitulate the full dynamics, but *does* provide an effective, and nontrivial, theoretical constraint. We hypothesize that the “consistent” regime ($\rho_{samp} \in (0, \rho_{theo})$, 3,856 entries) represents the degradation of correlations due to technical noise in the sequencing process and stochastic intermediates. The “inconsistent” regime ($\rho_{samp} \in (\rho_{theo}, 1)$, 279 entries) may stem from model misidentification, and could be explained by coupling between splicing events, resulting in a burst model closer to Eq. 26. Some of these apparently inconsistent correlations may also be due to the small sample sizes, as discussed in section S5.1. Finally, the “negative” regime ($\rho_{samp} < 0$, 750 entries) technically meets the constraint, but cannot actually be reproduced by the model. This does not appear to be an artifact of sample sizes, as evident from the strong negative peak in Figure S8 a. Instead, we speculate that enrichment in negative correlations is the signature of a more complicated regulatory schema that preferentially synthesizes some isoforms to the exclusion of others, rather than choosing

the splicing pathway randomly (as encoded in the derivation of Eq. 33).

We can exploit the analogous intergene model, encoded in Eq. 29, to try to predict the gene-gene correlation matrix (Fig. 1 e) based solely on the marginals, supposing *all* 500 highest-expressed genes fire simultaneously as a limiting case. For each gene, we consider the highest-abundance root transcript that can be fit by a negative binomial distribution, and identify its marginal burst size and efflux rate. Plugging these parameter estimates into Eq. 26, we obtain theoretical correlations ρ_{theo} and reconstruct the correlation matrix (Fig. 1 f). Finally, we compare the intragene sample correlations ρ_{samp} to the theoretical values in Fig. 1 g. These results represent the 119,805 nontrivial correlation matrix entries based on the 490 genes with well-fit roots. The theoretical constraint (sample correlation equal to or lower than predicted correlation) was met in 119,503 cases (99.7%).

Yet again, the model provides a nontrivial bound. We hypothesize that the consistent regime ($\rho_{samp} \in (0, \rho_{theo})$, 117,542 entries) represents the degradation of correlations due to stochastic effects outside the model, much as before. The correlations in the inconsistent regime ($\rho_{samp} \in (\rho_{theo}, 1)$, 302 entries) lie very close to the identity line, so we hypothesize they are mostly explained by small sample sizes, as discussed in section S5.1. Finally, the “negative” regime ($\rho_{samp} < 0$, 1,961 entries) is rare and does not appear to produce a strong signal in the correlation distribution (Figure S8 b), so we expect they also emerge from small sample sizes.

CONCLUSIONS

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. Furthermore, we have derived the analytical expressions for the generating functions and demonstrated their existence. The following expression gives the general solution for the joint PGF G :

$$G(\mathbf{u}, t) = \mathbb{E}_{\Theta} \left[G^{(0)}(\mathbf{U}(\mathbf{u}, t)) \exp \left(\int_0^t \sum_{\ell=1}^n \sum_{q=1}^{\omega} k_{\ell,q}(s) (M_{\ell,q}(\mathbf{U}(\mathbf{u}, s), s) - 1) ds \right) \right]. \quad (53)$$

This expression is modular with respect to \mathbf{U} , the set of characteristics defining the splicing and degradation network, $M_{\ell,q}$, the joint generating functions governing the bursting dynamics for all possible cotranscribing modules, $G^{(0)}$, the initial condition, and f_{θ} , the law governing the parameter distributions. The burst frequency and distribu-

tion can be time dependent, and describe deterministic driving by a latent process or regulator. By iteratively applying the dependence on the initial condition, we can write down analogous expressions for certain cell-cycle processes. The integral in Eq. 53 cannot be solved analytically for any but the simplest $M_{\ell,q}$. However, the form *guarantees* that the dynamics of a general system can be solved using quadrature and do not require matrix-based methods.

Special cases of Eq. 53 *can* be solved to provide useful insights into model complexity necessary to capture the summary statistics of living cells. To attempt to describe high gene-gene correlations, we have investigated several models for synchronized transcription, and found 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 (6) and high absolute gene-gene correlations (2,23) observed in living cells.

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 500 genes, we found that the constraints were met for 95.3% of the intragene transcript-transcript correlations and 99.7% of the intergene transcript-transcript correlations. Nevertheless, the model was insufficient to recapitulate the precise quantitative details, suggesting that more detailed biophysical models of regulated splicing and technical noise are necessary.

The formal computational complexity of this procedure is $O(\mathcal{N} \log \mathcal{N})$ in state space size. However, this complexity ostensibly arises from the inverse Fourier transform. In the practical regime (with a state space up to approximately 10^3), the scaling is sub-linear; unfortunately, full joint distributions are essentially intractable due to the “curse of dimensionality”—the *space* complexity of holding an array of size \mathcal{N} in memory (potentially requiring petabytes of storage even for small splicing networks, as in section S4.1) for the inverse Fourier transform. Nevertheless, this method can compute

marginals without having to solve the entire joint system, which would be intractable with matrix-based methods.

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 DAG coupled to a source and a set of sinks can be described using the current mathematical formalism. Furthermore, 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 (74,75).

Finally, we briefly touch upon the class of delay chemical master equations, and survey several recent advances in the field in section S6. Due to the non-Markovian nature of delayed systems, general probabilistic solutions are rare (76) 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.

CODE AVAILABILITY

Google Colab Python notebooks that reproduce the analyses and benchmarking are available at https://github.com/pachterlab/GP_2021_2.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2022.02.004>.

AUTHOR CONTRIBUTIONS

L.P. and G.G. designed the research and wrote the article. G.G. derived and implemented the analytical solutions, validated them against simulations, and performed the sequencing data analysis.

ACKNOWLEDGMENTS

The spectral solution was derived by Meichen Fang. G.G. and L.P. were partially funded by NIH U19MH114830. The DNA illustration used in Figs. S1, S2, and S4, modified from (70), is a derivative of the DNA Twe-moji 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.

SUPPORTING CITATIONS

References (77–85) appear in the supporting material.

REFERENCES

1. La Manno, G., R. Soldatov, ..., P. V. Kharchenko. 2018. RNA velocity of single cells. *Nature*. 560:494–498. <http://www.nature.com/articles/s41586-018-0414-6>.
2. Shalek, A. K., R. Satija, ..., A. Regev. 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. 498:236–240. <http://www.nature.com/articles/nature12172>.
3. Svensson, V., R. Vento-Tormo, and S. A. Teichmann. 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13:599–604. <http://www.nature.com/articles/nprot.2017.149>.
4. Peccoud, J., and B. Ycard. 1995. Markovian modeling of gene product synthesis. *Theor. Popul. Biol.* 48:222–234. <https://linkinghub.elsevier.com/retrieve/pii/S0040580985710271>.
5. Gardiner, C. 2004. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 3rd. Springer, pp. 145–147.
6. Dar, R. D., B. S. Razoooky, ..., L. S. Weinberger. 2012. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U S A.* 109:17454–17459. <http://www.pnas.org/cgi/doi/10.1073/pnas.1213530109>.
7. Sanchez, A., and I. Golding. 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science*. 342:1188–1193. <http://www.sciencemag.org/cgi/doi/10.1126/science.1242975>.
8. Bokes, P. 2021. Exact and WKB-approximate distributions in a gene expression model with feedback in burst frequency, burst size, and protein stability. *Discrete Continuous Dynamical Syst. - B* 0:0. <https://www.aims-science.org/article/doi/10.3934/dcdsb.2021126>.
9. Sugár, I., and I. Simon. 2014. Self-regulating genes. exact steady state solution by using Poisson representation. *Open Phys.* 12.9:615–627. <http://www.degruyter.com/view/j/phys.2014.12.issue-9/s11534-014-0497-0/s11534-014-0497-0.xml>.
10. M Stoeckius, M., C. Hafemeister, and P. Smibert. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods*. 14:865–868. <https://www.nature.com/articles/nmeth.4380>.
11. Peterson, V. M., K. X. Zhang, ..., J. A. Klappenbach. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35:936–939. <http://www.nature.com/articles/nbt.3973>.
12. Chung, H., C. N. Parkhurst, ..., A. Regev. 2021. Joint single-cell measurements of nuclear proteins and RNA *in vivo*. *Nat. Methods*. 18:1204–1212. <https://www.nature.com/articles/s41592-021-01278-1>.
13. Zheng, G. X. Y., J. M. Terry, ..., J. H. Bielas. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049. <http://www.nature.com/articles/ncomms14049>.
14. Qiu, P. 2020. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* 11:1169. <http://www.nature.com/articles/s41467-020-14976-9>.
15. Sonesson, C., A. Srivastava, ..., M. B. Stadler. 2021. Preprocessing choices affect RNA velocity results for droplet scRNA-seq data. *PLoS Comput. Biol.* 17:e1008585. <https://dx.plos.org/10.1371/journal.pcbi.1008585>.
16. Gorin, G., and L. Pachter. 2021. Length biases in single-cell RNA sequencing of pre-mRNA. *Preprint at bioRxiv*, <https://www.biorxiv.org/content/10.1101/2021.07.30.454514v1>.
17. Amrhein, L., K. Harsha, and C. Fuchs. 2019. A mechanistic model for the negative binomial distribution of single-cell mRNA counts, Preprint at *bioRxiv* <http://biorxiv.org/lookup/doi/10.1101/657619>.
18. Neuert, G., B. Munsky, ..., A. van Oudenaarden. 2013. Systematic identification of signal-activated stochastic gene regulation. *Science*. 339:584–587. <http://www.sciencemag.org/lookup/doi/10.1126/science.1231456>.
19. Jia, C., and R. Grima. 2020. Dynamical phase diagram of an auto-regulating gene in fast switching conditions. *J. Chem. Phys.* 152:174110. <http://aip.scitation.org/doi/10.1063/5.0007221>.
20. Huang, L., Z. Yuan, ..., T. Zhou. 2014. Feedback-induced counterintuitive correlations of gene expression noise with bursting kinetics. *Phys. Rev. E.* 90:052702. <https://link.aps.org/doi/10.1103/PhysRevE.90.052702>.
21. Pratapa, A., A. P. Jalihal, ..., T. M. Murali. 2020. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods*. 17:147–154. <http://www.nature.com/articles/s41592-019-0690-6>.
22. Ezer, D., V. Moignard, ..., B. Adryan. 2016. Determining physical mechanisms of gene expression regulation from single cell gene expression data. *PLoS Comput. Biol.* 12:e1005072. <http://dx.plos.org/10.1371/journal.pcbi.1005072>.
23. Iacono, G., R. Massoni-Badosa, and H. Heyn. 2019. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.*

- 20:110. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1713-4>.
24. Ham, L., R. D. Brackston, and M. P. H. Stumpf. 2020. Extrinsic noise and heavy-tailed laws in gene expression. *Phys. Rev. Lett.* 124:108101. <https://link.aps.org/doi/10.1103/PhysRevLett.124.108101>.
 25. Beentjes, C. H. L., R. Perez-Carrasco, and R. Grima. 2020. Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. *Phys. Rev. E.* 101:032403. <https://link.aps.org/doi/10.1103/PhysRevE.101.032403>.
 26. Kazeev, V., M. Khammash, ..., C. Schwab. 2014. Direct solution of the chemical master equation using quantized tensor trains. *PLoS Comput. Biol.* 10:e1003359. <https://dx.plos.org/10.1371/journal.pcbi.1003359>.
 27. Kazeev, V., and C. Schwab. 2015. Tensor approximation of stationary distributions of chemical reaction networks. *SIAM J. Matrix Anal. Appl.* 36:1221–1247. <http://epubs.siam.org/doi/10.1137/130927218>.
 28. Sunkara, V. 2019. On the properties of the reaction counts chemical master equation. *Entropy.* 21:607. <https://www.mdpi.com/1099-4300/21/6/607>.
 29. Cont, R., and P. Tankov. 2004. *In Financial Modelling With Jump Processes.* Chapman & Hall/CRC.
 30. Singh, A., and P. Bokes. 2012. Consequences of mRNA transport on stochastic variability in protein levels. *Biophysical J.* 103:1087–1096. <https://linkinghub.elsevier.com/retrieve/pii/S0006349512007904>.
 31. West, D. B. 2001. *Introduction to Graph Theory, 2nd.* Prentice Hall, Upper Saddle River.
 32. Bondy, J. A., and U. S. R. Murty. 2008. *Graph Theory.* Springer London <https://link.springer.com/book/9781846289699>.
 33. Bokes, P., J. R. King, M. Loose, ..., 2012. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J. Math. Biol.* 64:829–854. <http://link.springer.com/10.1007/s00285-011-0433-5>.
 34. Jahnke, T., and W. Huisinga. 2006. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.* 54:1–26. <http://link.springer.com/10.1007/s00285-006-0034-x>.
 35. Gardiner, C. W., and S. Chaturvedi. 1977. The Poisson representation. I. A new technique for chemical master equations. *J. Stat. Phys.* 17:429–468. <http://link.springer.com/10.1007/BF01014349>.
 36. Iyer-Biswas, S., and C. Jayaprakash. 2014. Mixed Poisson distributions in exact solutions of stochastic auto-regulation models. *Phys. Rev. E.* 90:052712. <https://pubmed.ncbi.nlm.nih.gov/25493821/>.
 37. Iyer-Biswas, S., F. Hayot, and C. Jayaprakash. 2009. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E.* 79:031911. <https://link.aps.org/doi/10.1103/PhysRevE.79.031911>.
 38. Friedman, N., L. Cai, and X. S. Xie. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* 97:168302. <https://link.aps.org/doi/10.1103/PhysRevLett.97.168302>.
 39. Bokes, P. 2021. Heavy-tailed distributions in a stochastic gene autoregulation model. *J. Stat. Mech. Theor. Exp.* 2021:113403. <https://iopscience.iop.org/article/10.1088/1742-5468/ac2edb>.
 40. Jia, C., M. Q. Zhang, and H. Qian. 2017. Emergent Levy behavior in single-cell stochastic gene expression. *Phys. Rev. E.* 96:040402. <https://link.aps.org/doi/10.1103/PhysRevE.96.040402>.
 41. Karlis, D., and E. Kekalaki. 2005. Mixed Poisson distributions. *Int. Stat. Rev./Revue Internationale de Statistique.* 73:35–58. <http://www.jstor.org/stable/25472639>.
 42. Panjer, H. H. 2004–. Mixed Poisson Distributions. *In Encyclopedia of Actuarial Science.* J. L. Teugels, B. Sundt, and G. E. Willmot, eds. John Wiley & Sons, Ltd <https://onlinelibrary.wiley.com/doi/10.1002/9780470012505.tam022>.
 43. Barndorff-Nielsen, O. E., and N. Shephard. 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B.* 63:167–241. <https://rss.onlinelibrary.wiley.com/doi/10.1111/1467-9868.00282>.
 44. Petroni, N. C., and P. Sabino. 2020. Gamma Related Ornstein-Uhlenbeck Processes and their Simulation, Preprint at *arXiv* <http://arxiv.org/abs/2003.08810>.
 45. Barndorff-Nielsen, O. E., and N. Shephard. 2003. Integrated OU processes and non-Gaussian OU-based stochastic volatility models. *Scand. J. Stat.* 30:277–295. <https://www.jstor.org/stable/4616764>.
 46. Papadopoulos, N., P. R. Gonzalo, and J. Söding. 2019. PROSSTT: probabilistic simulation of single-cell RNA-seq data for complex differentiation processes. *Bioinformatics.* 35:3517–3519. <https://academic.oup.com/bioinformatics/article/35/18/3517/5305637>.
 47. Gorin, G., J. J. Vastola, M. Fang, and L. Pachter. 2021. Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments, Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.09.06.459173v1>.
 48. John, F. 1978. *Partial Differential Equations.* Springer US <https://link.springer.com/book/10.1007/978-1-4684-0059-5>.
 49. Gans, P. J. 1960. Open first-order stochastic processes. *J. Chem. Phys.* 33:691–694. <http://aip.scitation.org/doi/10.1063/1.1731239>.
 50. Golding, I., J. Paulsson, ..., E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123:1025–1036. <https://linkinghub.elsevier.com/retrieve/pii/S0092867405010378>.
 51. Cannoodt, R., W. Saelens, Y. Saeys, ..., 2021. Spearheading future omics analyses using *dyngen*, a multi-modal simulator of single cells. *Nat. Commun.* 12:3942. <http://www.nature.com/articles/s41467-021-24152-2>.
 52. Xu, H., L. A. Sepúlveda, ..., I. Golding. 2015. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nat. Methods.* 12:739–742. <http://www.nature.com/articles/nmeth.3446>.
 53. Huang, P., E. D. Pleasance, S. J. Jones, ..., 2007. Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. *Genome Res.* 17:1478–1485. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1987351/>.
 54. Moran, P. 1969. Statistical inference with bivariate gamma distributions. *Biometrika.* 56:627–634. <https://academic.oup.com/biomet/article-abstract/56/3/627/233798>.
 55. Yue, S., T. Ouarda, and B. Bobée. 2001. A review of bivariate gamma distributions for hydrological application. *J. Hydrol.* 246:1–18. <https://linkinghub.elsevier.com/retrieve/pii/S0022169401003742>.
 56. Blitzstein, J. K., and J. Hwang. 2015. *Introduction to probability. Texts in statistical science.* CRC Press, Taylor & Francis Group.
 57. Lindsay, B. G. 1995. Mixture models: theory, geometry and applications. *NSF-CBMS Reg. Conf. Ser. Probab. Stat.* 5:i–163. <https://www.jstor.org/stable/4153184>.
 58. Singer, Z., J. Yong, ..., M. Elowitz. 2014. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell.* 55:319–331. <https://linkinghub.elsevier.com/retrieve/pii/S1097276514005632>.
 59. Ham, L., D. Schnoerr, ..., M. P. H. Stumpf. 2020. Exactly solvable models of stochastic gene expression. *J. Chem. Phys.* 152:144106. <http://aip.scitation.org/doi/10.1063/1.5143540>.
 60. Sun, X.-M., A. Bowman, ..., S. Marguerat. 2020. Size-dependent increase in RNA polymerase II initiation rates mediates gene expression scaling with cell size. *Curr. Biol.* 30:1217–1230.e7. <https://linkinghub.elsevier.com/retrieve/pii/S096098222030097X>.
 61. Padovan-Merhar, O., G. Nair, ..., A. Raj. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell.* 58:339–352. <https://linkinghub.elsevier.com/retrieve/pii/S1097276515001707>.
 62. Cao, Z., and R. Grima. 2020. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proc. Natl. Acad. Sci. U S A.* 117:4682–4692. <http://www.pnas.org/lookup/doi/10.1073/pnas.1910888117>.
 63. Dubey, S. D. 1970. Compound gamma, beta and F distributions. *Metrika.* 16:27–31. <http://link.springer.com/10.1007/BF02613934>.

64. Pham-Gia, T., and Q. Duong. 1989. The generalized beta- and F-distributions in statistical modelling. *Math. Computer Model.* 12:1613–1625. <https://linkinghub.elsevier.com/retrieve/pii/0895717789903373>.
65. Milo, R., and R. Phillips. 2015. Cell Biology by the Numbers. *Garland Science* <https://www.taylorfrancis.com/books/mono/10.1201/9780429258770/cell-biology-numbers-ron-milo-rob-phillips>.
66. Skinner, S. O., H. Xu, ..., I. Golding. 2016. Single-cell analysis of transcription kinetics across the cell cycle. *eLife*. 5:e12175. <https://elifesciences.org/articles/12175>.
67. Dattani, J., and M. Barahona. 2017. Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *J. R. Soc. Interf.* 14:20160833. <https://royalsocietypublishing.org/doi/10.1098/rsif.2016.0833>.
68. Briggs, J. A., C. Weinreb, ..., A. M. Klein. 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*. 360:eaar5780. <http://www.sciencemag.org/lookup/doi/10.1126/science.aar5780>.
69. Zeisel, A., W. J. Kostler, ..., E. Domany. 2011. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.* 7:529. <http://msb.embopress.org/cgi/doi/10.1038/msb.2011.62>.
70. Gorin, G., and L. Pachter. 2020. Special function methods for bursty models of transcription. *Phys. Rev. E*. 102:022409. <https://link.aps.org/doi/10.1103/PhysRevE.102.022409>.
71. Perez-Carrasco, R., C. Beentjes, and R. Grima. 2020. Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance. *J. R. Soc. Interf.* 17:20200360. <https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0360>.
72. Zhou, T., and J. Zhang. 2012. Analytical results for a multistate gene model. *SIAM J. Appl. Mathematics*. 72:789–818. <http://epubs.siam.org/doi/10.1137/110852887>.
73. Tian, L., J. S. Jabbari, M. E. Ritchie, ..., 2020. Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biol.* 22:310. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02525-6>.
74. Xu, H., S. O. Skinner, ..., I. Golding. 2016. Stochastic kinetics of nascent RNA. *Phys. Rev. Lett.* 117:128101. <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.117.128101>.
75. Gorin, G., M. Wang, ..., H. Xu. 2020. Stochastic simulation and statistical inference platform for visualization and estimation of transcriptional kinetics. *PLoS One*. 15:e0230736. <https://dx.plos.org/10.1371/journal.pone.0230736>.
76. Leier, A., and T. T. Marquez-Lago. 2015. Delay chemical master equation: direct and closed-form solutions. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* 471:20150049. <https://royalsocietypublishing.org/doi/10.1098/rspa.2015.0049>.
77. MacDonald, N. 1978. Time Lags in Biological Models. Springer-Verlag, Berlin, Heidelberg <http://link.springer.com/10.1007/978-3-642-93107-9>.
78. Burrage, K., T. Tian, and P. Burrage. 2004. A multi-scaled approach for simulating chemical reaction systems. *Prog. Biophys. Mol. Biol.* 85:217–234. <https://linkinghub.elsevier.com/retrieve/pii/S007961070400029X>.
79. Gedeon, T., and P. Bokes. 2012. Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations. *Biophysical J.* 103:377–385. <https://linkinghub.elsevier.com/retrieve/pii/S0006349512006820>.
80. Miekisz, J., J. Poleszczuk, ..., U. Foryś. 2011. Stochastic models of gene expression with delayed degradation. *Bull. Math. Biol.* 73:2231–2247. <http://link.springer.com/10.1007/s11538-010-9622-4>.
81. Fatehi, F., Y. N. Kyrchko, and K. B. Blyuss. 2020. A new approach to simulating stochastic delayed systems. *Math. Biosciences*. 322:108327. <https://linkinghub.elsevier.com/retrieve/pii/S002556420300225>.
82. Barrio, M., K. Burrage, ..., T. Tian. 2006. Oscillatory regulation of *hes1*: discrete stochastic delay modelling and simulation. *PLoS Comput. Biol.* 2:e117. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0020117>.
83. Lafuerza, L. F., and R. Toral. 2011. Exact solution of a stochastic protein dynamics model with delayed degradation. *Phys. Rev. E*. 84:051121. <https://journals.aps.org/pre/abstract/10.1103/PhysRevE.84.051121>.
84. Lafuerza, L. F., and R. Toral. 2011. Role of delay in the stochastic creation process. *Phys. Rev. E*. 84:021128. <https://journals.aps.org/pre/abstract/10.1103/PhysRevE.84.021128>.
85. Jia, T., and R. V. Kulkarni. 2011. Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. *Phys. Rev. Lett.* 106:058102. <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.106.058102>.