Transient and delay chemical master equations

Gennady Gorin[1], Shawn Yoshida[1,2], and Lior Pachter[2,3,*]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125
[2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, 91125
[3]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125
[*]Correspondence: lpachter@caltech.edu

Abstract The serial nature of reactions involved in the RNA life-cycle motivates the incorporation of delays in models of transcriptional dynamics. The models couple a bursty or switching promoter to a fairly general set of Markovian or deterministically delayed monomolecular RNA interconversion reactions with no feedback. We provide numerical solutions for the RNA copy number distributions the models induce, and solve several systems with splicing and degradation. An analysis of single-cell and single-nucleus RNA sequencing data using these models reveals that the kinetics of nuclear export do not appear to require invocation of a non-Markovian waiting time.

Keywords Chemical master equations, delay processes, single-cell RNA sequencing, model identification, stochastic simulation

Acknowledgments G.G. thanks Dr. John J. Vastola and Catherine Felce for valuable discussions. G.G. and L.P. were partially funded by the National Institutes of Health grant U19MH114830. A part of the reported results were obtained during a Data Sciences Co-op with Celsius Therapeutics, Inc. The DNA and RNA illustrations are derived from the DNA Twemoji by Twitter, Inc., used under CC-BY 4.0.

Code availability All scripts used to process data and generate the figures are located at https://github.com/pachterlab/GYP_2022. The raw count matrices, as well as the outputs of the *Monod* pipeline, are available on Zenodo (Gorin, 2022).

Statements and Declarations The authors declare no competing interests.

1 INTRODUCTION

In spite of their substantial successes over the past few decades, Markov models may not be sufficient to describe biological processes. For example, consider the simplified RNA life-cycle depicted in Figure 1a: the unspliced RNA is transcribed from a DNA template, spliced within the nucleus, exported to the cytoplasm, and eventually degraded. There is considerable evidence that bacterial and mammalian transcription is typically Markovian and bursty (Dar et al., 2012, Fukaya et al., 2016, Lammers et al., 2020, Rodriguez and Larson, 2020, Sanchez and Golding, 2013), i.e., transcriptional events are apparently uniformly distributed in time, and generate multiple mRNA in rapid succession. This behavior is usually explained by appealing to a Markovian multi-state promoter model (Peccoud and Ycard, 1995); genes that exhibit more complex transcriptional dynamics are modeled by more sophisticated, but analogous Markovian models of transitions between promoter states (Ham et al., 2020b, Lammers et al., 2020, Zhou and Zhang, 2012). Similarly, the results of genome-wide inhibition experiments are largely consistent with Markovian RNA degradation, with exponentially decaying RNA levels over time (Sharova et al., 2009).

On the other hand, the kinetics of splicing and export are rather less well-characterized. For computational and mathematical tractability, our previous studies (Gorin and Pachter, 2021, 2022b) have elided the fine details of these processes. Thus, we typically assume that splicing is a single-step Markovian process, whereas export is rapid enough to neglect. These assumptions appear sufficient (Gorin and Pachter, 2021, 2022b) to fit data generated by single-cell RNA sequencing (scRNA-seq) that can quantify spliced (Zheng et al., 2017) and unspliced (La Manno et al., 2018, Melsted et al., 2021) RNA in whole cells (Figure 1b, gold and black arrows), but they have not been studied in detail.

The growth of single-cell RNA sequencing has been accompanied by advances in single-nucleus RNA sequencing (snRNA-seq) (Ding et al., 2020, Habib et al., 2017), which quantifies RNA in nuclei (Figure 1b, blue and black arrows). This technology does not provide a readout of cytoplasmic RNA, which occasionally reduces its utility for identifying genes of interest (Thrupp et al., 2020); however, it simplifies the experimental tissue handling, as scRNA-seq is incompatible with samples that are frozen or challenging to dissociate (Litviňuková et al., 2020, Selewa et al., 2020). The appropriate models for the genome-wide analysis of snRNA-seq data are not yet clear. We may plausibly retain the assumption that transcriptional and splicing dynamics are bursty and Markovian, consistently with single-cell data. However, we can no longer make the assumption of rapid export. As shown in Figure 1c, nuclear data are considerably enriched in unspliced RNA, but spliced RNA are present and the process of their export from the observed nuclear system must be mechanistically parameterized to fit the bivariate molecule distributions.

It is not *a priori* obvious that this export process should be Markovian. Previous theoretical studies have adopted this assumption (Filatova et al., 2022, Singh and Bokes, 2012), and some experimental studies involving synthetic fluorescent reporter constructs (Hansen et al., 2018, Munsky et al., 2018) have not observed gross discrepancies with the Markovian hypothesis. However, in previous work (Gorin and Pachter, 2022a), we observed that scRNA-seq data were largely consistent with the simple CME model of bursty transcription, Markovian splicing, and Markovian removal of spliced molecules, but snRNA-seq data exhibited a high rate of apparent model misspecification according to a goodness-of-fit criterion. We speculated that this phenomenon could be explained by a qualitative difference between mechanisms observable in single-nucleus and single-cell data. In single-cell systems, the spliced "efflux" step may reflect the relatively slow, Markovian degradation of spliced RNA, whereas in single-nucleus systems, it may reflect a more rapid, non-Markovian nuclear export process. However, the discrepancy may also be attributable to hard-to-model technical noise contributions (such as background counts introduced by the lysis process, or incompletely lysed cells erroneously sequenced along with nuclei), the violation of model assumptions (such as internal homogeneity), or the arbitrary choice of thresholds for the chi-squared test, necessitating dedicated model comparisons.

In the current study, we evaluate the consistency of the Markovian hypothesis in a more controlled fashion. We solve the models with deterministically delayed splicing and export, fit them for thousands of genes across five cell types from two species, and compare them to the standard Markovian model, using paired single-nucleus and single-cell datasets. We seek to answer the following three questions, ordered from least to most general:

- Do single-nucleus data *require* models that are substantially different from single-cell data, i.e., is the Markovian efflux hypothesis valid in nuclei?
- Given that transcriptional elongation is strongly non-Markovian (Corrigan et al., 2016), and splicing often (Coté et al., 2021) occurs co-transcriptionally (Drexler et al., 2020), is the Markovian splicing hypothesis consistent with sequencing data in general?
- To what extent do single-cell and single-nuclei data allow model identification?

Although the modeling, inference, and model discrimination procedures are quantitative, the answers we seek are largely qualitative. Furthermore, the strength of the conclusions is necessarily limited by the usual assumptions made for tractability, including the omission of technical noise, cell cycle, cell size effects, uncertainty in cell type assignments, and the granular

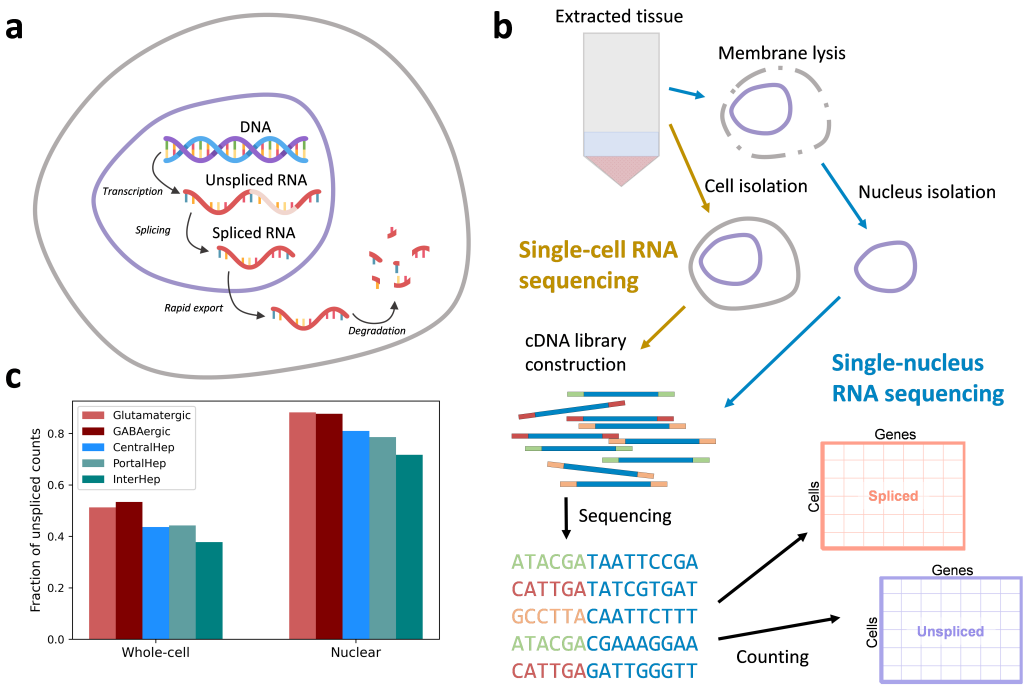

Figure 1 An outline of the experimental differences between single-cell and single-nucleus sequencing technologies. **a.** A schematic of our proposed model of the RNA life-cycle (gray outline: cell membrane; violet outline: nuclear membrane). **b.** The experimental workflows: single-cell sequencing uses whole cells, whereas single-nucleus sequencing uses isolated nuclei; both protocols produce a set of cDNA polymers that are sequenced and quantified to produce cell-by-gene matrices for the spliced and unspliced species. **c.** The differences in the experimental protocols translate to substantial and consistent differences in relative amounts of spliced and unspliced RNA: in single-cell datasets, unspliced RNA comprise 40–50% of the total dataset molecule counts, whereas in single-nucleus datasets, they predominate and comprise 70-90% of the total counts (colors: cell types; red: Allen mouse brain data; blue-green: Andrews human liver data). Illustrations adapted from (Gorin and Pachter, 2020b, Haque et al., 2017)

structure of distinct intermediate unspliced transcripts. For simplicity, we do not attempt to treat the full scope of possible non-Markovian dynamics, and focus on deterministically-timed delays.

The schema in Figure 1a, as well as the considerations in the previous section, motivate a bursty promoter coupled to a two-stage model of RNA processing for nuclear and whole-cell data; as usual, we adopt the assumption of rapid efflux for scRNA-seq. To solve these models under the Markovian or non-Markovian hypotheses, we treat a considerably more general problem with an arbitrary number of downstream species and upstream promoter states, which potentially reduce to a set of co-bursting modules.

Extensive literature exists on the topic of stochastic delay systems. In particular, the stepwise elongation of a nascent mRNA strand during the process of transcription is frequently described as a deterministic delay (in the nomenclature of queuing theory, a $G/D/\infty$ system) (Ali and Choubey, 2019, Choubey, 2018, Choubey et al., 2015, Fu et al., 2021, Xu et al., 2016, Zhang et al., 2014) or a series of small, stochastic steps along the template DNA strand (Gorin et al., 2020). Delayed regulation has been implicated in the p53 (Liu et al., 2011) and Hes1 systems (Barrio et al., 2006). Delay has been described as a regulatory motif in its own right, capable of modulating, suppressing, and buffering noise (Bhadana et al., 2021, Gedeon and Bokes, 2012, Singh et al., 2021, Zhang and Zhou, 2019), even beyond Markovian limits (Bhadana et al., 2021, Singh et al., 2021), as well as inducing oscillations (Bratsun et al., 2015, Liu et al., 2011, Miekisz et al., 2011, Singh et al., 2021).

In spite of these theoretical and biological applications, the studies described above generally discuss and derive statistical moments rather than full probability distributions, and few full analytical solutions are available for fitting transcriptomic copy number distributions. Several simple classes of delay chemical master equations (DCMEs) are solvable by the "linear chain trick" (LCT), which approximates a single constant delay of duration τ by $q \to \infty$ exponentially-distributed delays with rate $(\tau q)^{-1}$ (MacDonald, 1978). At least in principle, the same idea can be used for any Erlang delay distribution (Jia and Kulkarni, 2011, Lafuerza and Toral, 2011b, Leier and Marquez-Lago, 2015, Zhang and Zhou, 2019). Unfortunately, the LCT is

only practical when the distributions of all "virtual" intermediates are known in closed form, which is not the case for many biological systems of interest, such as bursty transcription (Gorin and Pachter, 2022a).

Several other analytical and numerical approaches have been proposed to solve DCMEs. There exists considerable literature on simulating delayed processes (Cai, 2007, Fatehi et al., 2020, Gupta et al., 2014, Tian et al., 2007), as well as approximating them by stochastic delay differential equations (SDDEs) (Gupta et al., 2014, Tian et al., 2007), which are typically more tractable. A recent report describes learning DCME solutions by designing a time-inhomogeneous CME with the time dependence of reaction rates encoded by a neural network (Jiang et al., 2021). Finally, reports by Lafuerza and Toral (Lafuerza and Toral, 2011a,b), Zhang and Zhou (Zhang and Zhou, 2019, Zhou and Zhang, 2012), Jiang et al. (Jiang et al., 2021), and Xu et al. (Xu et al., 2016) exploit the relationship between the DCME and closely related CMEs to obtain full probability distributions for several small systems.

In the current report, we find that this relationship is more generic and tractable than previously explored in the literature. We provide analytical and numerical recipes to determine the distributions of non-Markovian models with no feedback, and report and validate solutions for models with transcription governed by bursty and multi-state processes.

2 THEORETICAL FRAMEWORK

2.1 Preliminaries

We seek to model a system with n RNA species with counts $m_1, ..., m_n := \mathbf{m} \in \mathbb{N}_0^n$. The molecules interconvert through monomolecular reactions, which are either Markovian, with waiting time distributed per $\text{Exp}(c_{ik})$ for conversion from species i to species k, or delayed, with deterministic waiting time τ_i. The molecules may also be degraded, with analogous allowed waiting time distributions. We assume the interconversion topology is described by a directed acyclic graph (Gorin and Pachter, 2022a).

We consider two models of transcriptional dynamics. In the first, we suppose the transcription process is entirely memoryless, and molecules are produced in bursts according to a Poisson arrival process with rate α_i. This system is fully specified by a probability law $P(\mathbf{m}, t|P^0, 0)$, with the initial condition defined by the distribution $P^0(\mathbf{m})$.

In the second, we propose the transcribing promoter exists in one of N states, with interconversion rates H_{jk} for transitions from state j to state k. Each state has a characteristic transcription rate $\alpha_{j,i}$, where j is the state index and i is the molecule index; transcription occurs according to a Poisson arrival process. This system is fully specified by a probability law $\mathbf{P}(\mathbf{m}, t|\mathbf{w}, \mathbf{P}^0, 0)$, where the entries of $\mathbf{P}$ contain the non-normalized probabilities $P_j(\mathbf{m}, t)$ for each promoter state. The initial conditions include $\mathbf{P}^0$, the set of non-normalized molecule distributions for each state imposed at $t = 0$.

Solutions for these systems typically cannot be obtained in closed form. However, they can be calculated by the Fourier inversion of the probability generating function (PGF), defined as follows:

$$G(\mathbf{x}) = \sum_{\mathbf{m}} \mathbf{x}^{\mathbf{m}} P(\mathbf{m}), \tag{1}$$

where the summation takes place over all microstates $\mathbf{m} \in \mathbb{N}_0^n$, and $\mathbf{x}^{\mathbf{m}}$ is an informal shorthand for $\prod_i x_i^{m_i}$. It is typically more straightforward to consider $G(\mathbf{u})$, with $u_i = x_i - 1$. Further, the multi-state problem affords a PGF-like vector object $\mathbf{G}$, with components G_j such that

$$G_j(\mathbf{x}) = \sum_{\mathbf{m}} \mathbf{x}^{\mathbf{m}} P_j(\mathbf{m}). \tag{2}$$

2.2 The transient dynamics of bursty systems

As previously discussed in (Gorin and Pachter, 2022a), G^h, the PGF of a Markovian bursty process started at the homogeneous initial condition $\mathbf{m}_0 = 0$ has the following logarithm:

$$\ln G^h(\mathbf{u}, t) = \phi^h(\mathbf{u}, t) = \int_0^t \boldsymbol{\alpha}^T \left[\mathbf{M}(\mathbf{U}(\mathbf{u}, s)) - \mathbf{1}\right] ds, \tag{3}$$

where the inner product is taken over all distinct combinations of burst processes in the system. The entries of $\boldsymbol{\alpha}$ are the burst process arrival frequencies. The entries of $\mathbf{M}$ are the burst process moment-generating functions (MGFs). $\mathbf{u}$ is a vector of generating function arguments. The interconversion and degradation processes induce a series of *characteristics* $\mathbf{U}(\mathbf{u}, s)$, which solve the following series of ordinary differential equations:

$$\frac{d\mathbf{U}}{ds} = C\mathbf{U} \text{ s.t. } \mathbf{U}(s = 0) = \mathbf{u}, \tag{4}$$

where $C \in \mathbb{R}_+^{n\times n}$ is a matrix containing the rates of interconversion and degradation of the n species.

To obtain the PGF of a non-homogeneous system, started at a distribution with the arbitrary PGF $G^0(\mathbf{u})$, we take advantage of the fact that the newly transcribed molecules are statistically independent from the pre-existing molecules:

$$G(\mathbf{u}, t; G^0, 0) = G^0(\mathbf{U}(\mathbf{u}, t))G^h(\mathbf{u}, t). \tag{5}$$

This formula can be evaluated using a single integral per value of $\mathbf{u}$.

2.3 The transient dynamics of multi-state systems

It is straightforward to extend the foundations in (Gorin and Pachter, 2022a) to include interconversion between promoter states. We assume there are n species, with microstates $\mathbf{m} \in \mathbb{N}_0^n$, and N promoter states, with microstates $j \in \{0, ..., N-1\}$. The species are one-indexed; a reaction from a species to "0" corresponds to degradation. The promoter states are zero-indexed, to reflect the common notation wherein a "0" state is inactive whereas the "1" state is active.

We assume the system contains the following Markovian reactions:

- Interconversion between promoter states.
- Synthesis of new molecules.
- Interconversion of molecular species.
- Degradation of molecular species.

Define the collection of state-specific generating functions $\mathbf{G} := [G_0, ..., G_{N-1}]^T$, such that

$$G_j := \sum_{\mathbf{m}} \mathbf{x}^{\mathbf{m}} P_j(\mathbf{m}). \tag{6}$$

This collection is associated with the $N \times n$ Jacobian matrix, such that

$$J_{ji} := \frac{\partial G_j}{\partial x_i} = \frac{\partial G_j}{\partial u_i}, \tag{7}$$

defining $u_i = x_i - 1$ for each generating function argument.

By definition, the master equation terms for switching between states contribute the following terms to the PDE system:

$$\left(\frac{\partial G_j}{\partial t}\right)_{\text{switch}} = \sum_{j=0}^{N-1} H_{kj} G_k. \tag{8}$$

$H \in \mathbb{R}^N$ is the matrix encoding the continuous-time Markov chain (CTMC) governing the promoter states, such that $\sum_k H_{jk} = 0$ for all j. The diagonal terms encode the net efflux rate from state j, such that $H_{jj} := \sum_{k\neq j} H_{jk}$. This set of reactions can be represented in the usual form for a finite CTMC:

$$\left(\frac{\partial \mathbf{G}}{\partial t}\right)_{\text{switch}} = H^T \mathbf{G}. \tag{9}$$

From (Gorin and Pachter, 2022a), the master equation terms for RNA synthesis contribute the following terms to the PDE system:

$$\left(\frac{\partial G_j}{\partial t}\right)_{\text{tx}} = \sum_{i=1}^{n} \alpha_{j,i} u_i G_j, \tag{10}$$

This set of reactions can be summarized by the matrix $A \in (\mathbb{R}_{\geq 0})^{N\times n}$, such that $A_{ji} = \alpha_{j,i}$ is the rate of production of transcript i while the gene is in state j:

$$\left(\frac{\partial \mathbf{G}}{\partial t}\right)_{\text{tx}} = (A\mathbf{u}) \odot \mathbf{G}, \tag{11}$$

where the symbol $\odot$ denotes the Hadamard, or entrywise, product of two matrices of identical dimensions.

Finally, the master equations for monomolecular, non-catalytic interconversion or degradation contribute the following terms to the PDE system:

$$\left(\frac{\partial G_j}{\partial t}\right)_{\text{conv}} = \sum_{i=1}^{n}\left(-c_{i0}u_i + \sum_{k=1}^{n} c_{ik}(u_k - u_i)\right)\frac{\partial G_j}{\partial u_i} \tag{12}$$

where c_{ik} is the rate of transitioning from species i to species k and c_{i0} is the degradation rate of species i. This set of reactions can be summarized by the matrix $C \in \mathbb{R}^{n\times n}$ encoding the transitions between molecular species, such that $C_{ik} = c_{ik}$ and $C_{ii} = -\sum_{k\neq i} c_{ik} - c_{i0}$:

$$\left(\frac{\partial \mathbf{G}}{\partial t}\right)_{\text{conv}} = J(C\mathbf{u}) \tag{13}$$

The coupled generating function PDEs take the following form:

$$\begin{aligned} \frac{\partial \mathbf{G}}{\partial t} &= H^T\mathbf{G} + (A\mathbf{u}) \odot \mathbf{G} + J(C\mathbf{u}) \\ -H^T\mathbf{G} - (A\mathbf{u}) \odot \mathbf{G} &= -\frac{\partial \mathbf{G}}{\partial t} + J(C\mathbf{u}). \end{aligned} \tag{14}$$

For each j, we can apply the method of characteristics (Singh and Bokes, 2012):

$$\begin{aligned} \frac{dG_j}{ds} &= \frac{dG_j}{dT}\frac{dT}{ds} + \sum_{i=1}^{n}\frac{dG_j}{dU_i}\frac{dU_i}{ds} \\ \frac{dU_i}{ds} &= (C\mathbf{U})_i \\ \frac{dT}{ds} &= -1 \end{aligned} \tag{15}$$

Imposing $T(s = 0) = t$, we find $T(s) = t - s$. By enforcing $U_i(s = 0) = u_i$, we can simultaneously solve the system $\partial_s \mathbf{U} = C\mathbf{U}$. We then yield:

$$-\frac{d\mathbf{G}(\mathbf{U}(s), T(s))}{ds} = H^T\mathbf{G} + (A\mathbf{U}) \odot \mathbf{G}, \tag{16}$$

an initial value problem in N variables, parametrized by the characteristic variable s. We seek to obtain $\mathbf{G}$ at $T = t$, i.e., $s = 0$. We possess an initial condition $\mathbf{G}^0$ at $T = 0$, i.e., $s = t$. By numerically integrating $-H^T\mathbf{G} - (A\mathbf{U}) \odot \mathbf{G}$ from $s = t$ to $s = 0$, using $\mathbf{G}^0(\mathbf{U}(t))$ as the initial condition, we yield $\mathbf{G}(\mathbf{u}, t)$.

2.3.1 Solution through the lens of matrix ODEs

In practical terms, Equation 16 and its associated initial conditions are the solution to a particular switching system: the ODE can be easily plugged into a standard Runge–Kutta-type solver to obtain the generating function. However, we can, in principle, write down the ODE in more compact form:

$$\begin{aligned} \mathcal{D}(t) &= -H^T - \text{diag } A\mathbf{U} \\ \frac{d\mathbf{G}}{ds} &= \mathcal{D}(s)\mathbf{G}, \end{aligned} \tag{17}$$

where $\mathcal{D}$ is the matrix that encodes system *d*ynamics. We observe:

$$\mathcal{D}(t_1)\mathcal{D}(t_2) = \left[H^T + \text{diag } A\mathbf{U}(t_1)\right]\left[H^T + \text{diag } A\mathbf{U}(t_2)\right] \neq \mathcal{D}(t_2)\mathcal{D}(t_1) \tag{18}$$

in general; the product of diagonal matrices commutes, but the product of H^T and a diagonal matrix does not (except in the trivial case $N = 1$). Therefore, $\mathbf{G}$ cannot be represented by a finite matrix exponential, but can be approximated by a Magnus series (Iserles and MacNamara, 2017).

2.3.2 Solution through the lens of special functions

Certain steady-state solutions can be obtained by appealing to the theory of special functions. The stationary PGF of the solution to the standard telegraph model (with $N = 2$ and $n = 1$) is given by the confluent hypergeometric function ${}_1F_1$ (Iyer-Biswas et al., 2009); transient solutions have an analogous form. The combinatorial extension of the telegraph model (with $N = 2^p$, $p \in \mathbb{N}$), whose state transitions are given by the Hamming graph, has a stationary PGF given by the p-fold product of confluent hypergeometric functions (Ham et al., 2020b). Certain systems with $n = 1$ and an irreversible refractory state dynamics have solutions given by the generalized hypergeometric function ${}_{N-1}F_{N-1}$ (Zhou and Zhang, 2012). Thus, in principle, solutions may be obtained by casting the system of N ODEs in Equation 16 into a single order-N ODE, then solving it using the properties of special functions. However, this tends to be considerably more involved than the original problem, both due to the complexity of manipulating hypergeometric functions and the challenges of actually evaluating them. Further, although hypergeometric representations are available for $n = 1$, the case of $n > 1$ does not even afford a formal solution. We have found the ODE approach yields a satisfactory combination of numerical stability and robustness for most purposes.

2.4 Applications to delay master equations

The numerical recipes outlined above can be exploited to solve delay chemical master equations. Instead of using Equation 4 to obtain the characteristics, we solve the downstream system by hand, using Equations S35 and S36 of (Gorin and Pachter, 2022a) for species with Markovian efflux and the identity

$$U_i(s) = u_i \mathbb{I}(s < \tau_i) + U_k(s - \tau_i)\mathbb{I}(s > \tau_i) \tag{19}$$

for species i converted to species k after a deterministic delay τ_i. $\mathbb{I}$ denotes the indicator function, which returns $\mathbb{I}(A) = 1$ if A holds and 0 otherwise. If species i is degraded after a deterministic delay, its characteristic is simply $u_i \mathbb{I}(s < \tau_i)$. The resulting collection of characteristics can be plugged directly into Equation 5 or 16 and integrated.

3 RESULTS

3.1 Solving an arbitrary multi-state system

By numerical integration, we can obtain the molecule distributions induced by complex, arbitrary systems, such as the one shown in Figure 2a. To obtain the solution, we use Equation 16. We calculate $\mathbf{U}$, use $\mathbf{G}^0(\mathbf{U}(t))$ as the initial condition to numerically compute $\mathbf{G}$, and invert.

The parameters and initial conditions, given in Section 3.1.2, were set arbitrarily, taking care to keep the rates of interconversion reactions comparable and avoid regression to "trivial" regimes. The characteristics were computed using the MATLAB Symbolic Math Toolbox (MathWorks, 2022). As shown in Figure 2b, the procedure generates results that match simulations, and affords full conditional distributions at arbitrary times and for arbitrary initial conditions.

3.1.1 Simulation and solution procedure

To compute the analytical solutions, we applied a standard Runge-Kutta integration method, implemented using `integrate.rk45` in the *SciPy* package (Virtanen et al., 2020).

To simulate the delayed system, we implemented a custom algorithm that introduces three modifications to Gillespie's stochastic simulation algorithm (Gillespie, 1976). The Markovian form of the algorithm takes the following form:

1. Initialize system at time $t = t_0$ and state $x = x_0$.
2. Compute instantaneous reaction rates of the μth reaction, $\text{flux}_\mu(x)$ and the net state efflux rate, $\text{flux}(x) = \sum_\mu \text{flux}_\mu(x)$.
3. Compute sojourn time $t_s = \frac{1}{\text{flux}(x)} \log \frac{1}{u_1}$ and reaction index, μ, such that $\sum_{\nu=1}^{\mu-1} < u_2 \text{flux}(x) \leq \sum_{\nu=1}^{\mu}$, where u_1 and u_2 are random variables uniformly distributed over $[0, 1]$.
4. Advance system to time $t \leftarrow t + t_s$ and state $x \leftarrow x + \nu_\mu$.
5. Return to step 2 or terminate simulation.

The first modification treats removal events of delayed species, i.e, transcripts that undergo reactions with deterministic waiting times. Two empty queues are initialized: one for times and one for reaction indices. Then, if the reaction index determined in step 3 is the creation of a delayed species, the queues are populated with the time and reaction index of the deterministic removal of that species in the correct order. If the system is initialized with delayed species, the queues of times and reaction indices must be pre-defined accordingly. For simplicity, we always assume that existing delayed species were created at $t = 0$.

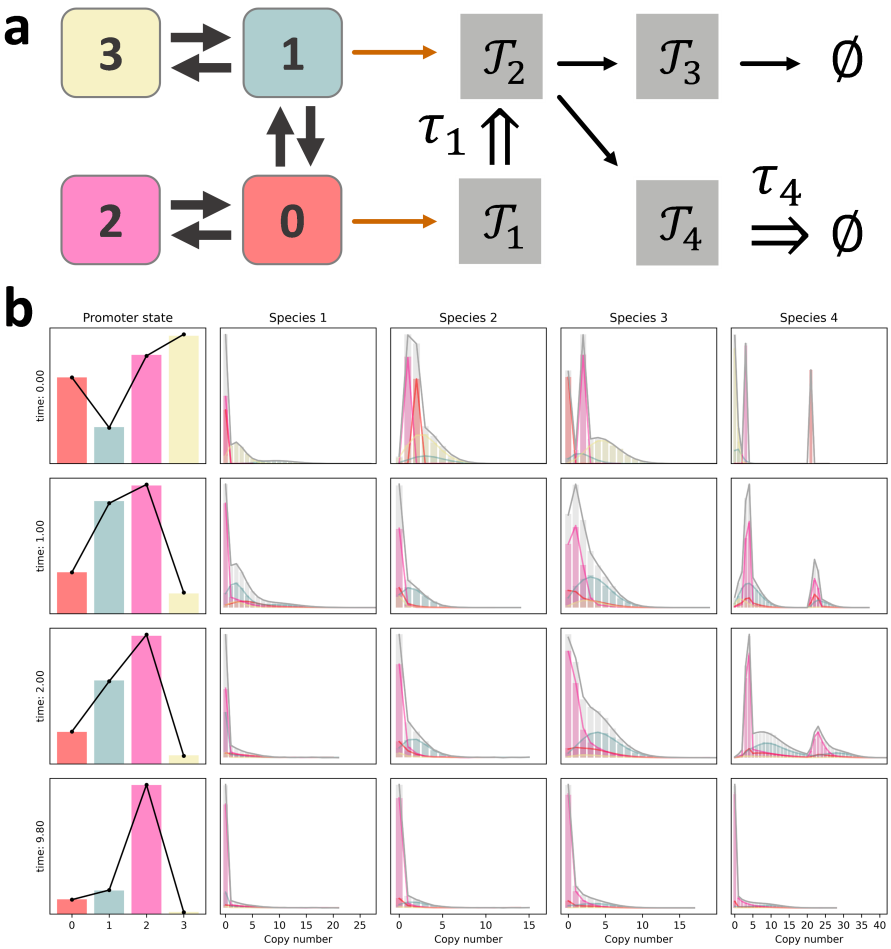

Figure 2 Validating a complicated non-Markovian transcriptional system. **a**. The structure of a system with four promoter states and four RNA species, two of which have deterministic lifetimes (rounded squares: promoter states; thick arrows: promoter state transitions; gray squares: RNA species; thin orange arrows: transcription reactions; thin black arrows: Markovian splicing and degradation reactions; wide arrows: deterministically delayed reactions). **b**. The analytical solutions for marginal distributions of promoter states and molecular species match simulations (Rows: time points. Leftmost column: promoter state distributions; histograms: simulation; lines: matrix exponential solution. Columns 2–5: the marginal distributions of molecular species; colored histograms: non-normalized marginals at each promoter state; gray histograms: molecular marginals aggregated over promoter states; lines: ODE solutions)

Table 1 Kinetic parameters used to simulate the multi-state DCME model

Parameter	Value
H_{01}	0.75
H_{10}	0.5
H_{02}	1
H_{20}	0.1
H_{13}	0.15
H_{31}	2
$\alpha_{0,1}$	5.2
$\alpha_{1,2}$	9.2
c_{23}	1.8
c_{24}	2.5
c_{30}	0.8
τ_1	1.1
τ_4	2.1

Table 2 Initial conditions used to simulate the multi-state DCME model

	State 0	State 1	State 2	State 3
Probability	0.24	0.1	0.3	0.36
Species 1	D-0	P-10	D-0	P-2.3
Species 2	D-2	P-3.5	D-1	P-3
Species 3	D-0	P-2	D-2	P-5
Species 4	D-21	P-1.1	D-3	P-0.2

The second modification alters the calculation of flux, specifically accounting for the contributions of species that don't yet exist, but will after some delay. The total flux, $\text{flux}(x)$, is computed at each queued reaction event, and sojourn time corresponding to the random flux generated by $\log \frac{1}{u_1}$ is found analytically. The computation of the sojourn time is essentially equivalent to the direct method outlined in (Cai, 2007).

The third modification ensures that all reactions happen in the correct order. After step 3, the reaction time and event are stored, and before advancing the system in step 4, all queued reactions that are to happen before the stored reaction event are sequentially applied and stored. The resulting ordered list is then converted into system times and states.

The algorithm is implemented in Python, and is designed to simulate the class of systems outlined in Section 2.3. With minor modifications, it can be used to simulate multi-molecular reactions, catalysis, feedback schema, or more general waiting time distributions, analogously to (Cai, 2007).

3.1.2 Simulation parameters and characteristic solutions

The parameters for the system 2a are given in Table 1. The global parameters are $N = 4$ and $n = 4$. H_{ij} denotes the rate of transition from promoter state i to promoter state j. $\alpha_{j,i}$ denotes the rate of production of transcript $\mathcal{T}_i$ at promoter state j. c_{ik} denotes the rate of conversion of $\mathcal{T}_i$ to $\mathcal{T}_k$. c_{i0} denotes the rate of degradation of $\mathcal{T}_i$. r_i denotes the total efflux rate $\sum_k c_{ik}$. τ_i denotes the residence time for deterministically delayed conversion. We generated realizations of trajectories for 10,000 cells, simulated until $t = 10$.

The initial conditions are outlined in Table 2. The first row reports the probability of the system starting in each state. The following rows encode the distributions of each species at $t = 0$, conditional on starting in each state. D-x corresponds to a deterministic initial condition with x molecules, whereas P-μ corresponds to a Poisson initial condition with an average of μ molecules.

The system is described by the following characteristics:

$$
\begin{aligned}
U_4 &= u_4 \mathbb{I}(s < \tau_4) \\
U_3 &= u_3 e^{-c_{30} s} \\
U_2 &= u_2 e^{-r_2 s} \\
&\quad + u_3 \frac{c_{23}}{r_2 - c_{30}} \left[e^{-c_{30} s} - e^{-r_2 s} \right] \\
&\quad + u_4 \frac{c_{24}}{r_2} \left[\mathbb{I}(s < \tau_4) + \mathbb{I}(s > \tau_4) e^{-r_2 (s - \tau_4)} - e^{-r_2 s} \right] \\
U_1 &= u_1 \mathbb{I}(s < \tau_1) \\
&\quad + u_2 \mathbb{I}(s > \tau_1) e^{-r_2 (s - \tau_1)} \\
&\quad + u_3 \frac{c_{23}}{r_2 - c_{30}} \mathbb{I}(s > \tau_1) \left[e^{-s (c_{30} - \tau_1)} - e^{-r_2 (s - \tau_1)} \right] \\
&\quad + u_4 \frac{c_{24}}{r_2} \left[\mathbb{I}(s \in (\tau_1, \tau_4)) - \mathbb{I}(s > \tau_1) e^{-r_2 (s - \tau_1)} + \mathbb{I}(s > \tau_1 + \tau_4) e^{-r_2 (s - \tau_1 - \tau_4)} \right]
\end{aligned}
\tag{20}
$$

3.2 Single-nucleus data do not show identifiable mechanistic differences from single-cell data

As discussed in Section 1, we do not *a priori* know that splicing and nuclear export are Markovian. Using the suite of mathematical tools described in Section 2.2 and relatively homogeneous data, we revisit (Gorin and Pachter, 2021) the question of RNA life-cycle modeling and consider the differences between nuclear and whole-cell systems in further detail.

We are primarily interested in comparing the stationary distributions induced by the DCME reaction system with delayed efflux

$$
\varnothing \xrightarrow{\alpha} B \times \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \overset{\tau}{\Longrightarrow} \varnothing \tag{21}
$$

to those induced by the CME system

$$
\varnothing \xrightarrow{\alpha} B \times \mathcal{T}_1 \xrightarrow{\beta} \mathcal{T}_2 \xrightarrow{\gamma} \varnothing, \tag{22}
$$

where B is distributed according to a geometric distribution on $\mathbb{N}_0$ with mean b. We solve the DCME system in Section 3.2.2. In addition, we solve the closely related system with delayed splicing, or deterministic retention of the molecule at the gene locus, in Section 3.2.3:

$$
\varnothing \xrightarrow{\alpha} B \times \mathcal{T}_1 \overset{\tau}{\Longrightarrow} \mathcal{T}_2 \xrightarrow{\gamma} \varnothing. \tag{23}
$$

By using a computer algebra system, such as the MATLAB Symbolic Math Toolbox (MathWorks, 2022), it is straightforward to obtain the lower moments. For comparison, Table 3 displays the moments for the Markovian and delayed systems.

Table 3 Lower moments of the models of interest

Moment	Markovian	Delayed efflux	Delayed splicing
μ_1	$\frac{\alpha b}{\beta}$	$\frac{\alpha b}{\beta}$	$\alpha b \tau$
σ_1^2	$\mu_1(1+b)$	$\mu_1(1+b)$	$\mu_1(1+2b)$
μ_2	$\frac{\alpha b}{\gamma}$	$\alpha b \tau$	$\frac{\alpha b}{\gamma}$
σ_2^2	$\mu_2 \left(1 + \frac{b\beta}{\beta+\gamma}\right)$	$\mu_2 + 2\alpha b^2 \left[\tau + \frac{1}{\beta}\left(e^{-\beta\tau} - 1\right)\right]$	$\mu_2(1+b)$
$\mathrm{Cov}(X_1, X_2)$	$\frac{\alpha b^2}{\beta+\gamma}$	$\frac{\alpha b^2}{\beta}\left[1 - e^{-\beta\tau}\right]$	0

We fit the three models introduced above, as well as a qualitatively similar overdispersed model with a gamma-distributed transcription rate (Gorin and Pachter, 2020a, Gorin et al., 2021, Ham et al., 2020a) (Figure 3a), interpreting unspliced RNA counts as $\mathcal{T}_1$ and spliced counts as $\mathcal{T}_2$. The analysis was performed using *Monod* (Gorin and Pachter, 2022b), using gradient descent to obtain maximum likelihood estimates (MLEs) under the four models (Section 3.2.4). As misspecification was evident

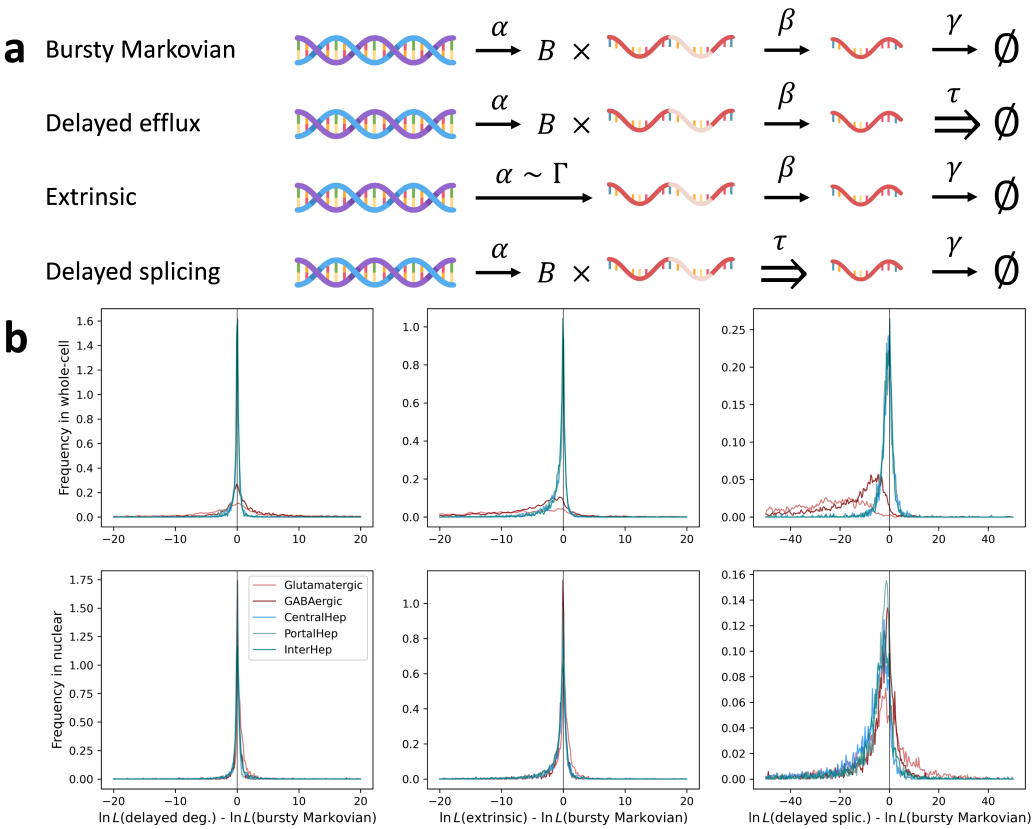

Figure 3 Certain models of RNA transcription and processing may be distinguished based on single-cell data, but single-nucleus data do not suggest that nuclear export biophysics are identifiably non-Markovian. **a.** The reaction schema of the considered models: DNA generates unspliced RNA with transcriptional burst frequency α, the unspliced RNA are converted to spliced RNA; the spliced RNA are removed from the system. In the "extrinsic" model, α is gamma-distributed; in the others, it is identical for all cells, and the transcriptional events produce bursts of size $B \sim$ Geom with mean b. **b.** Single-cell data (top row) support the bursty Markovian and delayed-efflux models to approximately the same degree, but offer consistently less support to the other two models, as quantified by the likelihood ratio. Single-nucleus data (bottom row) appear to show similar broad trends, but at lower magnitudes, limiting model identifiability (colors: cell types; red: Allen data; blue-green: Andrews data; lines: kernel density estimates)

with or without length bias in the original report (cf. `brain_5k_nuc` in Figs. S5 and S6 of (Gorin and Pachter, 2021)), we omitted the modeling of technical noise. The models were separately fit to GABAergic and glutamatergic cell types from two mouse brain samples (Yao et al., 2021), one generated using single-cell sequencing and one generated using single-nucleus sequencing, as well as single-cell and single-nucleus data from pericentral, periportal, and interzonal hepatocytes from a single human liver sample (Andrews et al., 2022).

Upon fitting the MLEs, we computed the models' likelihood ratios relative to the standard bursty Markovian model and plotted them (Figure 3b). We did not observe a systematic bias toward the delayed-efflux model in the nuclear data (left column); the log-likelihood ratios were near zero in all considered cases, suggesting the data were insufficient to strongly favor either model in any of the datasets, nuclear or whole-cell. In contrast, the other two models were less favored relative to the standard model, with log-likelihood ratios tending to be negative. The strength of evidence against the models was lower in the single-nucleus data. This loss of statistical identifiability concords with intuition: the bursty Markovian, delayed-efflux, and extrinsic models afford identical negative binomial unspliced RNA marginals, requiring a large amount of spliced RNA to distinguish the models, which the nuclear sequencing protocol lacks. The delayed-splicing model has a different nascent marginal (described by the geometric-Poisson or Pólya–Aeppli distribution), and appeared to be more distinguishable from the negative binomial.

3.2.1 Bursty Markovian system

The steady state of the system given in Equation 22 has been previously reported (Singh and Bokes, 2012) and has the log-PGF

$$\begin{aligned} U_1^M(s;\beta,\gamma) &= u_2\frac{\beta}{\beta-\gamma}e^{-\gamma s} + \left(u_1 - u_2\frac{\beta}{\beta-\gamma}\right)e^{-\beta s} \\ \phi^M &= \alpha\int_0^\infty \left[\frac{1}{1-bU_1^M(s;\beta,\gamma)} - 1\right] ds, \end{aligned} \tag{24}$$

where U_1^M is the characteristic of the Markovian system.

3.2.2 Bursty with delayed efflux

The system given in Equation 21 is governed by the following equations:

$$\begin{aligned} U_2^D &= u_2\mathbb{I}(s<\tau) \\ \frac{dU_1^D}{ds} &= \beta\left(U_2^D - U_1^D\right) \\ U_1^D(s;\beta,\tau) &= u_1e^{-\beta s} + u_2\left(1-e^{-\beta s}\right) - u_2\left(1-e^{-\beta(s-\tau)}\right)\mathbb{I}(s>\tau) \end{aligned} \tag{25}$$

where $\mathbb{I}$ is the indicator function. This can be converted to a piecewise constant representation:

$$\begin{aligned} U_1^D(s;\beta,\tau) &= u_2 + (u_1-u_2)e^{-\beta s} \text{ whenever } s\in(0,\tau) \text{ and} \\ &= (u_1-u_2)e^{-\beta s} + u_2e^{-\beta(s-\tau)} \\ &= \left(u_2 + (u_1-u_2)e^{-\beta\tau}\right)e^{\beta\tau}e^{-\beta s} \text{ whenever } s\in(\tau,\infty). \end{aligned} \tag{26}$$

Defining $U(s;\beta) := u_2 + (u_1-u_2)e^{-\beta s}$ for convenience and integrating:

$$\begin{aligned} \phi^D &= \alpha\int_0^\tau \left[\frac{1}{1-bU_1^D(s;\beta,\tau)} - 1\right] ds + \alpha\int_\tau^\infty \left[\frac{1}{1-bU_1^D(s;\beta,\tau)} - 1\right] ds \\ &= \frac{\alpha/\beta}{1-bu_2}\ln\left(\frac{bU(\tau;\beta)-1}{bu_1-1}\right) + \alpha\tau\frac{bu_2}{1-bu_2} - \frac{\alpha}{\beta}\ln\left(1-bU(\tau;\beta)\right) \end{aligned} \tag{27}$$

3.2.3 Bursty with delayed splicing

The system given in Equation 23 supposes that each molecule of nascent RNA is retained in the system for duration τ. This is model is typical for studies concerned with modeling submolecular details of transcriptional elongation (Xu et al., 2016). This system is governed by the following characteristics:

$$\begin{aligned} U_2^D &= u_2e^{-\gamma s} \\ U_1^D(s;\tau,\gamma) &= u_1\mathbb{I}(s<\tau) + u_2e^{-\gamma(s-\tau)}\mathbb{I}(s>\tau) \end{aligned} \tag{28}$$

Upon plugging the characteristics into the burst size MGF and integrating, we find:

$$\begin{aligned} \phi^D &= \alpha\int_0^\tau \left[\frac{1}{1-bU_1^D(s;\gamma,\tau)} - 1\right] ds + \alpha\int_\tau^\infty \left[\frac{1}{1-bU_1^D(s;\gamma,\tau)} - 1\right] ds \\ &= \frac{\alpha\tau bu_1}{1-bu_1} - \frac{\alpha}{\gamma}\ln(1-bu_2), \end{aligned} \tag{29}$$

i.e., the joint distribution is merely the product of two independent marginal distributions. This result can be derived by appealing to the bursty limit of the argument in Section 14.4 in the supplementary material of (Xu et al., 2016).

Table 4 Sources and cell counts of analyzed cell type datasets

Source	Reference	Donor	Dataset	Cell type	Cells
Allen	Yao et al. (2021)	457911	B08	Glutamatergic	5352
Allen	Yao et al. (2021)	457911	B08	GABAergic	789
Allen	Yao et al. (2021)	427311	A02	Glutamatergic	6731
Allen	Yao et al. (2021)	427311	A02	GABAergic	1253
Andrews	Andrews et al. (2022)	4	C72 sc	CentralHep	1567
Andrews	Andrews et al. (2022)	4	C72 sc	PortalHep	4421
Andrews	Andrews et al. (2022)	4	C72 sc	InterHep	991
Andrews	Andrews et al. (2022)	4	C72 sn	CentralHep	2885
Andrews	Andrews et al. (2022)	4	C72 sn	PortalHep	2862
Andrews	Andrews et al. (2022)	4	C72 sn	InterHep	1879

3.2.4 Data analysis

To compare the biophysical mechanisms consistent with single-cell and single-nucleus datasets, we used mouse cortex data generated by Allen Institute for Brain Science ("Allen data") (Yao et al., 2021) and human liver data generated by Toronto's University Health Network ("Andrews data") (Andrews et al., 2022). The datasets analyzed for this study are reported in Table 4.

The Allen data did not contain technical replicates, i.e., a single mouse tissue processed using two single-cell and single-nucleus technologies. In lieu of this, we chose two libraries from female mice: B08, obtained from donor 457911, and A02, obtained from donor 427311. The raw single-cell FASTQs were obtained from `https://data.nemoarchive.org/biccn/grant/u19_zeng/zeng/transcriptome/scell/10x_v3/mouse/raw/MOp/`; the annotations were obtained from `https://data.nemoarchive.org/biccn/grant/u19_zeng/zeng/transcriptome/scell/10x_v3/mouse/processed/analysis/10X_cells_v3_AIBS/`. The raw single-nucleus FASTQs were obtained from `https://data.nemoarchive.org/biccn/grant/u19_zeng/zeng/transcriptome/sncell/10x_v3/mouse/raw/MOp/`; the annotations were obtained from `https://data.nemoarchive.org/biccn/grant/u19_zeng/zeng/transcriptome/sncell/10x_v3/mouse/processed/analysis/10X_nuclei_v3_AIBS/`.

For quantification, we used a pre-built mm10 reference genome released by 10x Genomics, obtained from `https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest`, to generate cDNA and intron references using *kallisto*|*bustools* 0.26.0 (`kb ref` with option `--lamanno`). Both libraries were produced using the 10x Genomics v3 chemistry. We pseudoaligned the B08 and B02 libraries using the 10x v3 cell barcode whitelist and the default *kallisto*|*bustools* filter (`kb count` with options `--lamanno`, `-x 10xv3`, and `--filter bustools`) to generate spliced and unspliced count matrices. The data were further filtered, excluding all cells with too few total (spliced + unspliced) UMIs, with a threshold of 10^4 for B08 and 8×10^3 for A02. These thresholds were chosen manually based on knee plots.

To choose cells for analysis by *Monod*, we identified cell barcodes which were (1) assigned to either the GABAergic or glutamatergic cell type according to annotations, (2) passed the *kallisto*|*bustools* filter and knee plot filters, and (3) did not belong to one of the excluded rare cell subtypes (Scng or L6 IT Car3). This procedure yielded the cell numbers reported in the final column of Table 4.

The Andrews data contained technical replicates, generated using different sequencing chemistries. We analyzed data for donor 4 (`C72_reseq` for single-cell, here referred to as "C72 sc"; `C72_TST` for single-nucleus, here referred to as "C72 sn"). The single-cell data were obtained in the Sequence Read Archive format (experiment SRX12509405, runs SRR16227578–83) and converted to paired-end FASTQ files with *fasterq-dump* 2.11.2 (with options `--include-technical` and `-S`). The single-nucleus data were obtained in the originally submitted BAM format (experiment SRX12509406, run SRR16227584) and converted to paired-end FASTQ files with the 10x Genomics software *bamtofastq* 1.3.1. The metadata were obtained from the Gene Expression Omnibus (series GSE185477, file `GSE185477_Final_Metadata.txt.gz`).

For quantification, we used a pre-built GRCh38 reference genome released by 10x Genomics, obtained from `https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest`, to generate cDNA and intron references using *kallisto*|*bustools* 0.26.0 (`kb ref` with option `--lamanno`). The single-cell library was produced using the 10x v2 chemistry, whereas the single-nucleus library was produced using the v3 chemistry. To simplify the procedure of extracting and storing the count data, we used the lists of barcodes present in each sample's cell type annotations as the whitelists for pseudoalignment. The *kallisto*|*bustools* was run using the `--lamanno` option and the appropriate `-x` arguments for each dataset. Due to the whitelist definition, we did not use the default filter. Instead, we excluded all cells with too few total

UMIs, with a threshold of 8×10^2 for single-cell and 2×10^3 for single-nucleus, and too many total UMIs, with corresponding thresholds of 10^4 and 5×10^4. These thresholds were chosen manually based on knee plots.

To choose cells for analysis by *Monod*, we identified cell barcodes which were (1) assigned to pericentral (CentralHep), periportal (PortalHep), or interzonal (InterHep) hepatocyte cell types according to annotations and (2) passed the knee plot filters. This procedure yielded the cell numbers reported in the final column of Table 4.

The analysis was performed using *Monod* 0.2.4.0. For each species, we selected 3,000 genes with moderate to high expression in all datasets. This procedure imposed a set of constraints (each species' mean copy number > 0.01, each species' maximum copy number > 3 and < 400) and computed the number of datasets in which the criterion was met. The top 3,000 genes were identified by choosing the top quantiles of this metric, using random non-replacement sampling to break ties.

Next, we fit all of the models to the data. The procedure used gradient descent to minimize the Kullback-Leibler divergence (KLD) between model and data, and output the maximum likelihood estimates (MLEs) for the parameters. To compute the proposed probability distributions for the KLD, we evaluated the generating functions on a two-dimensional grid of microstates, with the lower bound set to zero and the upper bound set at 10 molecules more than the observed maximum of each species. The generating function of the bursty Markovian model, which requires numerical integration, was evaluated by order-60 Gaussian quadrature, implemented through `integrate.fixed_quad` in *SciPy* (Virtanen et al., 2020), evaluated on the interval $[0, T]$, where $T = 10\left(1 + \beta^{-1} + \gamma^{-1}\right)$ to ensure the system reached equilibrium. For the bursty models, α was set to unity with no loss of generality at steady state. For the extrinsic noise model, θ, the scale parameter of the gamma distribution, was set to unity.

Gradient descent was performed using L-BFGS-B, a constrained optimization algorithm, implemented through the *SciPy* function `optimize.minimize` (Virtanen et al., 2020). The optimization was performed in a 3-dimensional space of $\log_{10}$ parameter values, with domain of $(-2, 3.1)$ for the burst sizes and gamma shape parameters and $(-1.8, 3.5)$ for β, γ, and τ^{-1}. These parameters were chosen based on the parameter distributions seen in Figure S5 of (Gorin and Pachter, 2021). Optimization ran for a maximum of 15 steps, with 5 independent trials. The first trial was initialized at the method of moments (MoM) estimate. The method of moments estimates for the bursty Markovian models and the extrinsic models are reported in the supplement of (Gorin and Pachter, 2022b). The MoM estimate for the delayed efflux model was identical to that of the bursty Markovian model, with τ^{-1} taking the place of γ. The MoM estimate for the delayed splicing model was identical to that of the bursty Markovian model, with b corrected by a factor of $\frac{1}{2}$ and τ^{-1} taking the place of β. These estimates follow directly from the results in Table 3.

The log-likelihood ratios visualized in Figure 3 were evaluated by computing the proposal PMFs as above, then computing the difference between the log-likelihoods of the data under the alternative model and under the bursty model:

$$\ln \mathrm{LR}_{j,\mu} = \ln \mathcal{L}(\mathcal{D}_j | \hat{\Theta}_{j,\mu}) - \ln \mathcal{L}(\mathcal{D}_j | \hat{\Theta}_{j,\mathrm{bursty}}) \tag{30}$$

where μ is the alternative model type (delayed-efflux, extrinsic, delayed-splicing), j is the gene index, $\mathcal{D}_j$ is the dataset (spliced and unspliced copy numbers in a particular cell type) for gene j, $\hat{\Theta}_{j,\mu}$ is the vector of parameter MLEs under the model, and $\mathcal{L}$ is the data likelihood.

Log-likelihood ratios with magnitude > 100 for single-cell and > 50 for single-nucleus data were omitted as potential signatures of failure to converge to a satisfactory MLE. Kernel density estimates in Figure 3b were generated from the remaining $\ln \mathrm{LR}_j$ values using the SciPy class `stats.gaussian_kde`, with bandwidth set to 0.01 to avoid oversmoothing.

4 METHODOLOGICAL EXTENSIONS

In the current section, we draw useful connections to theory we previously reported in (Gorin and Pachter, 2021), contextualize the work with respect to standard tools of DCME analysis, and discuss potential extensions.

4.1 Generating function identifiability

Consider a system with n species, described by a generating function $G(\mathbf{u})$. Some of these species may not be mutually identifiable. Suppose there are m identifiable species, where each species $i \in \{1, ..., n\}$ may be assigned to each category $\imath$ with probability $p_{i,\imath}$ for each molecule. If this assignment process is independent, the process induces a collection of n generating functions $\mathbf{G}^s(\boldsymbol{u})$, such that each component is the PGF of a categorical distribution (i.e., a multinomial distribution with a single trial):

$$G_i^s(\boldsymbol{u}) = \sum_{\imath=0}^{m} (u_\imath + 1) p_{i,\imath}, \tag{31}$$

where $p_{i,0}$ is defined as $1 - \sum_i p_{i,i}$, the probability of assigning species i to none of the observable species, and losing it. Evidently,

$$G_i^s(\boldsymbol{u}) - 1 = \sum_{i=0}^{n} u_i p_{i,i}. \tag{32}$$

From standard properties of generating functions, the distribution resulting from "filtering" G through $\mathbf{G}^s$ is a function composition:

$$G(\boldsymbol{u}) = G(\mathbf{G}^s(\boldsymbol{u}) - \mathbf{1}), \tag{33}$$

where $\mathbf{1}$ denotes the length-n vector of ones.

4.2 Erlang-distributed delays

Consider the case of a system with bursty transcription, Markovian degradation, and Erlang-distributed waiting time for splicing, with shape q and mean waiting time τ. First, set up a system with $q + 1$ species:

$$\varnothing \xrightarrow{\alpha} B \times \mathcal{T}_1 \xrightarrow{q/\tau} ...\mathcal{T}_q \xrightarrow{q/\tau} \mathcal{T}_{q+1} \xrightarrow{\gamma} \varnothing \tag{34}$$

The characteristic corresponding to $\mathcal{T}_{q+1}$ is simply $u_{q+1}e^{-\gamma s}$. By somewhat tedious computation, amounting to repeatedly solving ordinary differential equations of the form $\frac{dU_i}{dt} = \frac{q}{\tau}(U_{i+1} - U_i)$, we find that the characteristic corresponding to $\mathcal{T}_1$ takes the following form:

$$U_1 = u_{q+1}e^{-\gamma s} \times \left(\frac{\lambda}{\lambda - \gamma}\right)^q + e^{-\lambda s}\sum_{i=0}^{q-1}\frac{(u_{i+1}\lambda s)^i}{i!} - u_{q+1}e^{-\lambda s}\frac{(\lambda)^q}{(\lambda-\gamma)^q}\sum_{i=0}^{q-1}\frac{s^i(\lambda-\gamma)^i}{i!}, \tag{35}$$

where λ is defined as q/τ.

To aggregate $\mathcal{T}_1, ..., \mathcal{T}_q$ into a single species $\mathcal{X}_1$, and represent $\mathcal{T}_{q+1}$ as the downstream species $\mathcal{X}_2$, we evaluate the arguments at $u_1 = ... = u_q = u_1$ and $u_{q+1} = u_2$. This yields:

$$U_1 = u_2 e^{-\gamma s} \times \left(\frac{\lambda}{\lambda - \gamma}\right)^q + u_1\frac{\Gamma(q, \lambda s)}{\Gamma(q)} - u_2 e^{-\gamma s}\frac{\lambda^q}{(\lambda-\gamma)^q}\frac{\Gamma\left(q, (\lambda-\gamma)s\right)}{\Gamma(q)}. \tag{36}$$

To evaluate the PGF, we plug U_1 into the burst distribution MGF and integrate.

4.3 Linear chain trick

The same approach can be used to arrive at the solution for deterministically delayed systems, commonly known as the "linear chain trick." Taking the limit as $q \to \infty$ and exploiting the properties of special functions:

$$\begin{aligned}\lim_{q\to\infty} U_1 &= u_2 e^{-\gamma s}e^{\gamma\tau} + u_1\mathbb{I}(s < \tau) - \lim_{q\to\infty} u_2 e^{-\gamma s}e^{\gamma\tau}\frac{\Gamma(q, (q/\tau - \gamma)s)}{\Gamma(q)} \\ &= u_2 e^{-\gamma(s-\tau)} + u_1\mathbb{I}(s < \tau) - u_2 e^{-\gamma(s-\tau)}\mathbb{I}(s < \tau) \\ &= u_1\mathbb{I}(s < \tau) + u_2 e^{-\gamma(s-\tau)}\mathbb{I}(s > \tau).\end{aligned} \tag{37}$$

To obtain the PGF, we plug $\lim_{q\to\infty} U_1$ into the burst distribution MGF and integrate, eliding the functional analysis justification for interchanging the limit and the integral. This reproduces the results in Section 3.2.3, albeit with considerably more effort.

4.4 General waiting time distributions

The same approach may be used to evaluate systems with more generic waiting time distributions. For example, the characteristic appropriate for a molecule that remains in the system for time τ, then exhibits Markovian efflux is immediately implied by the results of the previous section:

$$U = u\left[\mathbb{I}(s < \tau) + e^{-\gamma(s-\tau)}\mathbb{I}(s > \tau)\right], \tag{38}$$

obtained by evaluating the previous characteristic U_1 at $u_1 = u_2 = u$. This holds for any combination of Markovian and deterministic delays.

Throughout the current section, we have considered delays only in downstream molecular processing. However, studies have indicated that non-Markovian refractory periods can occur in promoter state transitions (Harper et al., 2011). If the refractory period can be effectively described by a series of Markovian processes (e.g., if the waiting time is Erlang), such systems can be solved by defining "internal" inactive states and adding their P_j, obtained through Equation 16, in the spirit of (Herbach, 2019). Stinchcombe et al. treated the case of arbitrary switching time distributions (Stinchcombe et al., 2012); however, this approach requires a degree of ingenuity and is fairly challenging to formalize in the language of generating functions.

The case of more general waiting times was explored in considerable detail by Zhang and Zhou (Zhang and Zhou, 2019), albeit with somewhat more emphasis on theoretical foundations and implications for noise buffering than here. Essentially, arbitrary waiting time distributions may be encoded by properly exploiting their statistical structure; for example, it is immediately evident that the characteristic in Equation 38 is simply the survival function (complementary cumulative distribution function) of each molecule, evaluated at its birth. In the same vein, the U_2^D terms in Equations 25 and 28 are the survival functions of the degenerate and exponential distributions, whereas the u_2 term of the characteristic in Equation 24 is the survival function of the two-parameter hypoexponential distribution.

It appears that more general waiting time distributions can be incorporated in the characteristic framework by appropriately manipulating their survival functions. We forgo detailed discussion of this extension, as it is less suitable to automation and tends to lead to intractable integrals. Further, Markovian and deterministic delays are relatively easy to motivate from first principles by appealing to memorylessness or perfect molecular memory, respectively (either maximum or zero entropy in the language of statistical physics (Pressé et al., 2013)). On the other hand, "intermediate" non-Markovian cases require a specific hypothesis to instantiate a functional form for the waiting time distribution.

The simulation algorithm outlined in Section 3.1.1 generalizes to arbitrary waiting time distributions, and may be used to investigate the impact of different assumptions even when analytical solutions are unavailable. Further, the tools outlined here can be used to explore some pathological cases, such as the loss of stationarity due to particularly ill-behaved waiting time distributions. For example, if the RNA production process is constitutive with unity transcription rate and the waiting time distribution is standard half-Cauchy, the characteristic and log-PGF take the following form:

$$\begin{aligned} U &= u\left[1 - \frac{2}{\pi}\tan^{-1} s\right] \\ \phi &= \frac{u}{\pi}\left[\pi t - 2t\tan^{-1}(t) + \ln(1+t^2)\right], \end{aligned} \tag{39}$$

where ϕ is immediately recognizable as the log-PGF of a Poisson distribution, albeit one that does not possess a limit as $t \to \infty$. This implies that the $M/G/\infty$ queue with half-Cauchy service times has no stationary distribution. In contrast, if the waiting time distribution is standard Pareto, with support on $[1, \infty)$, we find:

$$\begin{aligned} U &= us^{-\alpha} \\ \phi &= \frac{u}{\alpha - 1}\left(1 - t^{1-\alpha}\right) \\ \text{or } \phi &= u \ln t \text{ when } \alpha = 1. \end{aligned} \tag{40}$$

This distribution possesses a Poisson steady state with mean $(\alpha - 1)^{-1}$ whenever $\alpha > 1$. This implies that the $M/G/\infty$ queue with Pareto service times has a stationary distribution, but only if the service time distribution is endowed with a mean. Approaches such as this may be of use in identifying classes of non-Markovian processes admissible under particular axioms about biology, e.g., the existence of steady states, without assuming the existence of all Laplace transforms for the waiting time distributions (Zhang and Zhou, 2019).

5 DISCUSSION

This report extends the solutions reported in (Gorin and Pachter, 2022a) to a broad class of delayed systems, and unifies them with the theory of switching genes outlined in (Vastola, 2021). The approach augments the toolbox of stochastic biophysical analysis: it affords generating function-based analytical solutions for systems which would otherwise be outside the purview of the standard analysis of Markov chains. The derivation of these solutions does not require the usual manipulations of master equations at several time points, merely fairly standard integrals. These solutions may be fit (e.g., through the *Monod* implementation) and used as alternative hypotheses to investigate whether the standard assumptions of memorylessness hold, whether they are violated, or whether the available data are equivocal.

In Section 3.2, we used these analytical solutions to fit several superficially similar bivariate overdispersed models to single-cell and single-nucleus data, and observed that the typical Markovian model is a strong candidate, but cannot be easily distinguished from the competing delayed-efflux model, especially in the low-spliced expression single-nucleus data. On the other hand, two alternative models are consistently judged less plausible. This suggests that the assumption of Markovian, one-step splicing is useful despite its simplicity. This investigation is far from comprehensive. We forgo modeling technical noise or identifying cell types; using cell type definitions based on the spliced RNA counts introduces some degree of circularity. However, even with these limitations, the results illustrate fundamental challenges associated with single-nucleus data. To sum up, the data analysis is consistent with the following answers to the questions posed in Section 1:

- Single-nucleus data do not appear to require qualitatively different models. Both single-cell and single-nucleus data are consistent with the delayed efflux model.
- The Markovian splicing hypothesis is considerably more consistent with data than the delayed-splicing hypothesis.
- As may be expected, the low amounts of spliced RNA in nuclear data mean that only models with substantial differences in the unspliced distributions can be distinguished.

Furthermore, even when generating functions cannot be obtained, which is the typical case for systems with multiple RNA species, our method provides a simple numerical recipe for evaluating their distributions. The numerical approach is guaranteed to run in $O(N \ln N)$, can use off-the-shelf integration routines (e.g., quadrature for bursty systems and Runge–Kutta methods for switching systems), and enables the evaluation of arbitrary marginal distributions, which is not feasible for matrix- or simulation-based methods.

The approach is largely modular with respect to the specific details of the downstream processes, i.e., the causal relationships between molecules and the Markovian or deterministic waiting times for the reactions. In Section 4, we discuss a set of extensions to even more generic systems, with waiting times described by combinations of exponential and degenerate distributions. The same generating function-based approach can be used to implement models of technical noise and molecular non-identifiability under the assumption of independent sampling (Gorin and Pachter, 2021). Although the solutions are fairly generic, they do not yet provide a route to treating more complex systems with protein translation or molecular feedback. We speculate that a sufficiently general treatment of such systems may uncover similar relationships between CMEs and DCMEs, and provide a simulation routine that can be extended to model such phenomena.

REFERENCES

M. Z. Ali and S. Choubey. Decoding the grammar of transcriptional regulation from RNA polymerase measurements: models and their applications. *Physical Biology*, 16(6):061001, Oct. 2019. ISSN 1478-3975. doi: 10.1088/1478-3975/ab45bf. URL https://iopscience.iop.org/article/10.1088/1478-3975/ab45bf.

T. S. Andrews, J. Atif, J. C. Liu, C. T. Perciani, X.-Z. Ma, C. Thoeni, M. Slyper, G. Eraslan, A. Segerstolpe, J. Manuel, S. Chung, E. Winter, I. Cirlan, N. Khuu, S. Fischer, O. Rozenblatt-Rosen, A. Regev, I. D. McGilvray, G. D. Bader, and S. A. MacParland. Single-Cell, Single-Nucleus, and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal Heterogeneity. *Hepatology Communications*, 6(4):821–840, 2022. doi: https://doi.org/10.1002/hep4.1854. URL https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep4.1854.

M. Barrio, K. Burrage, A. Leier, and T. Tian. Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modelling and Simulation. *PLOS Computational Biology*, 2(9):e117, Sept. 2006. doi: 10.1371/journal.pcbi.0020117. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0020117.

J. Bhadana, A. L. Chanu, M. Z. Malik, and R. K. B. Singh. Noise and delay can shape distribution functions in stochastic reaction dynamics. *Nonlinear Dynamics*, 105(1):797–811, July 2021. ISSN 0924-090X, 1573-269X. doi: 10.1007/s11071-021-06643-5. URL https://link.springer.com/10.1007/s11071-021-06643-5.

D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty. Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences*, 102(41):14593–14598, Oct. 2015. doi: 10.1073/pnas.0503858102. URL https://www.pnas.org/doi/10.1073/pnas.0503858102.

X. Cai. Exact stochastic simulation of coupled chemical reactions with delays. *The Journal of Chemical Physics*, 126(12):124108, Mar. 2007. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.2710253. URL http://aip.scitation.org/doi/10.1063/1.2710253.

S. Choubey. Nascent RNA kinetics: Transient and steady state behavior of models of transcription. *Physical Review E*, 97(2): 022402, 2018. ISSN 2470-0045, 2470-0053. doi: 10.1103/PhysRevE.97.022402.

S. Choubey, J. Kondev, and A. Sanchez. Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. *PLOS Computational Biology*, 11(11):e1004345, 2015. ISSN 1553-7358.

A. M. Corrigan, E. Tunnacliffe, D. Cannon, and J. R. Chubb. A continuum model of transcriptional bursting. *eLife*, 5:e13051, Feb. 2016. ISSN 2050-084X. doi: 10.7554/eLife.13051. URL https://elifesciences.org/articles/13051.

A. Coté, C. Coté, S. Bayatpour, H. L. Drexler, K. A. Alexander, F. Chen, A. T. Wassie, E. S. Boyden, S. Berger, L. S. Churchman, and A. Raj. pre-mRNA spatial distributions suggest that splicing can occur post-transcriptionally. Preprint, bioRxiv: 2020.04.06.028092, June 2021. URL https://doi.org/10.1101/2020.04.06.028092.

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, Oct. 2012. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1213530109. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1213530109.

J. Ding, X. Adiconis, S. K. Simmons, M. S. Kowalczyk, C. C. Hession, N. D. Marjanovic, T. K. Hughes, M. H. Wadsworth, T. Burks, L. T. Nguyen, J. Y. H. Kwon, B. Barak, W. Ge, A. J. Kedaigle, S. Carroll, S. Li, N. Hacohen, O. Rozenblatt-Rosen, A. K. Shalek, A.-C. Villani, A. Regev, and J. Z. Levin. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature Biotechnology*, 38(6):737–746, June 2020. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-020-0465-8. URL https://www.nature.com/articles/s41587-020-0465-8.

H. L. Drexler, K. Choquet, and L. S. Churchman. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Molecular Cell*, 77(5):985–998.e8, Mar. 2020. ISSN 10972765. doi: 10.1016/j.molcel.2019.11.017. URL https://linkinghub.elsevier.com/retrieve/pii/S1097276519308652.

F. Fatehi, Y. N. Kyrychko, and K. B. Blyuss. A new approach to simulating stochastic delayed systems. *Mathematical Biosciences*, 322:108327, Apr. 2020. ISSN 00255564. doi: 10.1016/j.mbs.2020.108327. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556420300225.

T. Filatova, N. Popović, and R. Grima. Modulation of nuclear and cytoplasmic mRNA fluctuations by time-dependent stimuli: Analytical distributions. *Mathematical Biosciences*, 347:108828, May 2022. ISSN 00255564. doi: 10.1016/j.mbs.2022.108828. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556422000372.

X. Fu, H. P. Patel, S. Coppola, L. Xu, Z. Cao, T. L. Lenstra, and R. Grima. Accurate inference of stochastic gene expression from nascent transcript heterogeneity. Preprint, bioRxiv: 2021.11.09.467882, Nov. 2021. URL http://biorxiv.org/lookup/doi/10.1101/2021.11.09.467882.

T. Fukaya, B. Lim, and M. Levine. Enhancer Control of Transcriptional Bursting. *Cell*, 166(2):358–368, July 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.05.025. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867416305736.

T. Gedeon and P. Bokes. Delayed Protein Synthesis Reduces the Correlation between mRNA and Protein Fluctuations. *Biophysical Journal*, 103(3):377–385, Aug. 2012. ISSN 00063495. doi: 10.1016/j.bpj.2012.06.025. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349512006820.

D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, Dec. 1976. ISSN 00219991. doi: 10.1016/0021-9991(76)90041-3. URL https://linkinghub.elsevier.com/retrieve/pii/0021999176900413.

G. Gorin. Supporting data for GYP_2022 (10.5281/zenodo.7217195), Oct. 2022. URL https://zenodo.org/record/7217195.

G. Gorin and L. Pachter. Intrinsic and extrinsic noise are distinguishable in a synthesis – export – degradation model of mRNA production. Preprint, bioRxiv: 2020.09.25.312868, Sept. 2020a. URL http://biorxiv.org/lookup/doi/10.1101/2020.09.25.312868.

G. Gorin and L. Pachter. Special function methods for bursty models of transcription. *Physical Review E*, 102(2):022409, Aug. 2020b. ISSN 2470-0045, 2470-0053. doi: 10.1103/PhysRevE.102.022409. URL https://link.aps.org/doi/10.1103/PhysRevE.102.022409.

G. Gorin and L. Pachter. Length Biases in Single-Cell RNA Sequencing of pre-mRNA. Preprint, bioRxiv: 2021.07.30.454514, July 2021. URL https://www.biorxiv.org/content/10.1101/2021.07.30.454514v1.

G. Gorin and L. Pachter. Modeling bursty transcription and splicing with the chemical master equation. *Biophysical Journal*, 121(6):1056–1069, Feb. 2022a. doi: 10.1016/j.bpj.2022.02.004. URL https://www.cell.com/biophysj/fulltext/S0006-3495(22)00104-7.

G. Gorin and L. Pachter. *Monod*: mechanistic analysis of single-cell RNA sequencing count data. Preprint, bioRxiv: 2022.06.11.495771, June 2022b. URL http://biorxiv.org/lookup/doi/10.1101/2022.06.11.495771.

G. Gorin, M. Wang, I. Golding, and H. Xu. Stochastic simulation and statistical inference platform for visualization and estimation of transcriptional kinetics. *PLOS ONE*, 15(3):e0230736, Mar. 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0230736. URL https://dx.plos.org/10.1371/journal.pone.0230736.

G. Gorin, J. J. Vastola, M. Fang, and L. Pachter. Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments. Preprint, bioRxiv: 2021.09.06.459173, Sept. 2021. URL https://www.biorxiv.org/content/10.1101/2021.09.06.459173v1.

C. Gupta, J. M. López, R. Azencott, M. R. Bennett, K. Josić, and W. Ott. Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations. *The Journal of Chemical Physics*, 140(20):204108, May 2014. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.4878662. URL http://aip.scitation.org/doi/10.1063/1.4878662.

N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, D. A. Weitz, O. Rozenblatt-Rosen, F. Zhang, and A. Regev. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods*, 14(10):955–958, Oct. 2017. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.4407. URL http://www.nature.com/articles/nmeth.4407.

L. Ham, R. D. Brackston, and M. P. H. Stumpf. Extrinsic Noise and Heavy-Tailed Laws in Gene Expression. *Physical Review Letters*, 124(10):108101, Mar. 2020a. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.124.108101. URL https://link.aps.org/doi/10.1103/PhysRevLett.124.108101.

L. Ham, D. Schnoerr, R. D. Brackston, and M. P. H. Stumpf. Exactly solvable models of stochastic gene expression. *The Journal of Chemical Physics*, 152(14):144106, Apr. 2020b. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.5143540. URL http://aip.scitation.org/doi/10.1063/1.5143540.

M. M. Hansen, R. V. Desai, M. L. Simpson, and L. S. Weinberger. Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell Systems*, 7(4):384–397.e6, Oct. 2018. ISSN 24054712. doi: 10.1016/j.cels.2018.08.002. URL https://linkinghub.elsevier.com/retrieve/pii/S240547121830317X.

A. Haque, J. Engel, S. A. Teichmann, and T. Lönnberg. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, 9(1):75, Dec. 2017. ISSN 1756-994X. doi: 10.1186/s13073-017-0467-4. URL http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-017-0467-4.

C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, D. G. Spiller, J. J. Mullins, D. A. Rand, J. R. E. Davis, and M. R. H. White. Dynamic Analysis of Stochastic Transcription Cycles. *PLoS Biology*, 9(4):e1000607, Apr. 2011. ISSN 1545-7885. doi: 10.1371/journal.pbio.1000607. URL https://dx.plos.org/10.1371/journal.pbio.1000607.

U. Herbach. Stochastic Gene Expression with a Multistate Promoter: Breaking Down Exact Distributions. *SIAM Journal on Applied Mathematics*, 79(3):1007–1029, Jan. 2019. ISSN 0036-1399, 1095-712X. doi: 10.1137/18M1181006. URL https://epubs.siam.org/doi/10.1137/18M1181006.

A. Iserles and S. MacNamara. Magnus expansions and pseudospectra of Master Equations. Preprint, arXiv: 1701.02522, Jan. 2017. URL http://arxiv.org/abs/1701.02522. arXiv: 1701.02522.

S. Iyer-Biswas, F. Hayot, and C. Jayaprakash. Stochasticity of gene products from transcriptional pulsing. *Physical Review E*, 79(3):031911, Mar. 2009. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.79.031911. URL https://link.aps.org/doi/10.1103/PhysRevE.79.031911.

T. Jia and R. V. Kulkarni. Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting. *Physical Review Letters*, 106:058102, Feb. 2011. doi: 10.1103/PhysRevLett.106.058102. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.106.058102.

Q. Jiang, X. Fu, S. Yan, R. Li, W. Du, Z. Cao, F. Qian, and R. Grima. Neural network aided approximation and parameter inference of non-Markovian models of gene expression. *Nature Communications*, 12(1):2618, Dec. 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22919-1. URL http://www.nature.com/articles/s41467-021-22919-1.

G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, and P. V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, Aug. 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0414-6. URL http://www.nature.com/articles/s41586-018-0414-6.

L. F. Lafuerza and R. Toral. Exact solution of a stochastic protein dynamics model with delayed degradation. *Physical Review E*, 84:051121, Nov. 2011a. doi: 10.1103/PhysRevE.84.051121. URL https://journals.aps.org/pre/abstract/10.1103/PhysRevE.84.051121.

L. F. Lafuerza and R. Toral. Role of delay in the stochastic creation process. *Physical Review E*, 84:021128, Aug. 2011b. doi: 10.1103/PhysRevE.84.021128. URL https://journals.aps.org/pre/abstract/10.1103/PhysRevE.84.021128.

N. C. Lammers, Y. J. Kim, J. Zhao, and H. G. Garcia. A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting. *Current Opinion in Cell Biology*, 67:147–157, Dec. 2020. ISSN 09550674. doi: 10.1016/j.ceb.2020.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0955067420300971.

A. Leier and T. 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. ISSN 1364-5021, 1471-2946. doi: 10.1098/rspa.2015.0049. URL https://royalsocietypublishing.org/doi/10.1098/rspa.2015.0049.

M. Litviňuková, C. Talavera-López, H. Maatz, D. Reichart, C. L. Worth, E. L. Lindberg, M. Kanda, K. Polanski, M. Heinig, M. Lee, E. R. Nadelmann, K. Roberts, L. Tuck, E. S. Fasouli, D. M. DeLaughter, B. McDonough, H. Wakimoto, J. M. Gorham, S. Samari, K. T. Mahbubani, K. Saeb-Parsy, G. Patone, J. J. Boyle, H. Zhang, H. Zhang, A. Viveiros, G. Y. Oudit, O. A. Bayraktar, J. G. Seidman, C. E. Seidman, M. Noseda, N. Hubner, and S. A. Teichmann. Cells of the adult human heart. *Nature*, 588(7838):466–472, Dec. 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2797-4. URL https://www.nature.com/articles/s41586-020-2797-4.

B. Liu, S. Yan, Q. Wang, and S. Liu. Oscillatory expression and variability in p53 regulatory network. *Physica D: Nonlinear Phenomena*, 240(3):259–264, Feb. 2011. ISSN 01672789. doi: 10.1016/j.physd.2010.09.004. URL https://linkinghub.elsevier.com/retrieve/pii/S016727891000254X.

N. MacDonald. *Time Lags in Biological Models*, volume 27 of *Lecture Notes in Biomathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1978. ISBN 978-3-540-09092-2 978-3-642-93107-9. doi: 10.1007/978-3-642-93107-9. URL http://link.springer.com/10.1007/978-3-642-93107-9.

T. MathWorks. MATLAB R2022a Symbolic Math Toolbox, 2022. URL https://www.mathworks.com/products/symbolic.html.

P. Melsted, A. S. Booeshaghi, L. Liu, F. Gao, L. Lu, K. H. Min, E. da Veiga Beltrame, K. E. Hjörleifsson, J. Gehring, and L. Pachter. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nature Biotechnology*, 39(7):813–818, July 2021. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-021-00870-2. URL http://www.nature.com/articles/s41587-021-00870-2.

J. Miekisz, J. Poleszczuk, M. Bodnar, and U. Foryś. Stochastic Models of Gene Expression with Delayed Degradation. *Bulletin of Mathematical Biology*, 73(9):2231–2247, Sept. 2011. ISSN 0092-8240, 1522-9602. doi: 10.1007/s11538-010-9622-4. URL http://link.springer.com/10.1007/s11538-010-9622-4.

B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert. Distribution shapes govern the discovery of predictive models for gene regulation. *Proceedings of the National Academy of Sciences*, 115(29):7533–7538, 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1804060115.

J. Peccoud and B. Ycard. Markovian Modeling of Gene Product Synthesis. *Theoretical Population Biology*, 48(2):222–234, 1995. doi: 10.1006/tpbi.1995.1027.

S. Pressé, K. Ghosh, J. Lee, and K. A. Dill. Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics*, 85(3):1115–1141, July 2013. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.85.1115. URL https://link.aps.org/doi/10.1103/RevModPhys.85.1115.

J. Rodriguez and D. R. Larson. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annual Review of Biochemistry*, 89(1):189–212, June 2020. ISSN 0066-4154, 1545-4509. doi: 10.1146/annurev-biochem-011520-105250. URL https://www.annualreviews.org/doi/10.1146/annurev-biochem-011520-105250.

A. Sanchez and I. Golding. Genetic Determinants and Cellular Constraints in Noisy Gene Expression. *Science*, 342(6163): 1188–1193, Dec. 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1242975. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1242975.

A. Selewa, R. Dohn, H. Eckart, S. Lozano, B. Xie, E. Gauchat, R. Elorbany, K. Rhodes, J. Burnett, Y. Gilad, S. Pott, and A. Basu. Systematic Comparison of High-throughput Single-Cell and Single-Nucleus Transcriptomes during Cardiomyocyte Differentiation. *Scientific Reports*, 10:1535, Jan. 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-58327-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6992778/.

L. V. Sharova, A. A. Sharov, T. Nedorezov, Y. Piao, N. Shaik, and M. S. Ko. Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. *DNA Research*, 16(1):45–58, Jan. 2009. ISSN 1340-2838, 1756-1663. doi: 10.1093/dnares/dsn030. URL https://academic.oup.com/dnaresearch/article-lookup/doi/10.1093/dnares/dsn030.

A. Singh and P. Bokes. Consequences of mRNA Transport on Stochastic Variability in Protein Levels. *Biophysical Journal*, 103(5):1087–1096, Sept. 2012. ISSN 00063495. doi: 10.1016/j.bpj.2012.07.015. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349512007904.

S. N. Singh, A. L. Chanu, M. Z. Malik, and R. B. Singh. Interplay of cellular states: Role of delay as control mechanism. *Physica A: Statistical Mechanics and its Applications*, 572:125869, June 2021. ISSN 03784371. doi: 10.1016/j.physa.2021.125869. URL https://linkinghub.elsevier.com/retrieve/pii/S0378437121001412.

A. R. Stinchcombe, C. S. Peskin, and D. Tranchina. Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression. *Physical Review E*, 85(6):061919, June 2012. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.85.061919. URL https://link.aps.org/doi/10.1103/PhysRevE.85.061919.

N. Thrupp, C. Sala Frigerio, L. Wolfs, N. G. Skene, N. Fattorelli, S. Poovathingal, Y. Fourne, P. M. Matthews, T. Theys, R. Mancuso, B. de Strooper, and M. Fiers. Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Reports*, 32(13):108189, Sept. 2020. ISSN 22111247. doi: 10.1016/j.celrep.2020.108189. URL https://linkinghub.elsevier.com/retrieve/pii/S2211124720311785.

T. Tian, K. Burrage, P. M. Burrage, and M. Carletti. Stochastic delay differential equations for genetic regulatory networks. *Journal of Computational and Applied Mathematics*, 205(2):696–707, Aug. 2007. ISSN 03770427. doi: 10.1016/j.cam.2006.02.063. URL https://linkinghub.elsevier.com/retrieve/pii/S0377042706003943.

J. J. Vastola. *In search of a coherent theoretical framework for stochastic gene regulation*. PhD thesis, Vanderbilt, Mar. 2021. URL https://ir.vanderbilt.edu/handle/1803/16646.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders,

F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, Mar. 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0686-2. URL http://www.nature.com/articles/s41592-019-0686-2.

H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding. Stochastic Kinetics of Nascent RNA. *Physical Review Letters*, 117(12): 128101, 2016. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.117.128101. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.117.128101.

Z. Yao, H. Liu, F. Xie, S. Fischer, R. S. Adkins, A. I. Aldridge, S. A. Ament, A. Bartlett, M. M. Behrens, K. Van den Berge, D. Bertagnolli, H. R. de Bézieux, T. Biancalani, A. S. Booeshaghi, H. C. Bravo, T. Casper, C. Colantuoni, J. Crabtree, H. Creasy, K. Crichton, M. Crow, N. Dee, E. L. Dougherty, W. I. Doyle, S. Dudoit, R. Fang, V. Felix, O. Fong, M. Giglio, J. Goldy, M. Hawrylycz, B. R. Herb, R. Hertzano, X. Hou, Q. Hu, J. Kancherla, M. Kroll, K. Lathia, Y. E. Li, J. D. Lucero, C. Luo, A. Mahurkar, D. McMillen, N. M. Nadaf, J. R. Nery, T. N. Nguyen, S.-Y. Niu, V. Ntranos, J. Orvis, J. K. Osteen, T. Pham, A. Pinto-Duarte, O. Poirion, S. Preissl, E. Purdom, C. Rimorin, D. Risso, A. C. Rivkin, K. Smith, K. Street, J. Sulc, V. Svensson, M. Tieu, A. Torkelson, H. Tung, E. D. Vaishnav, C. R. Vanderburg, C. van Velthoven, X. Wang, O. R. White, Z. J. Huang, P. V. Kharchenko, L. Pachter, J. Ngai, A. Regev, B. Tasic, J. D. Welch, J. Gillis, E. Z. Macosko, B. Ren, J. R. Ecker, H. Zeng, and E. A. Mukamel. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature*, 598(7879):103–110, Oct. 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03500-8. URL https://www.nature.com/articles/s41586-021-03500-8.

J. Zhang and T. Zhou. Markovian approaches to modeling intracellular reaction processes with molecular memory. *Proceedings of the National Academy of Sciences*, 116(47):23542–23550, Nov. 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1913926116. URL https://pnas.org/doi/full/10.1073/pnas.1913926116.

X. Zhang, H. Jin, Z. Yang, and J. Lei. Effects of elongation delay in transcription dynamics. *Mathematical Biosciences and Engineering*, 11(6):1431–1448, 2014. ISSN 1551-0018. doi: 10.3934/mbe.2014.11.1431. URL http://www.aimspress.com/article/10.3934/mbe.2014.11.1431.

G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, and J. H. Bielas. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, Apr. 2017. ISSN 2041-1723. doi: 10.1038/ncomms14049. URL http://www.nature.com/articles/ncomms14049.

T. Zhou and J. Zhang. Analytical Results for a Multistate Gene Model. *SIAM Journal on Applied Mathematics*, 72(3):789–818, Jan. 2012. ISSN 0036-1399, 1095-712X. doi: 10.1137/110852887. URL http://epubs.siam.org/doi/10.1137/110852887.