bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# Direct simulation of a stochastically driven multi-step birth-death process

Gennady Gorin[1] and Lior Pachter[2]

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

March 4, 2021

## 1 Abstract

The description of transcription as a stochastic process provides a framework for the analysis of intrinsic and extrinsic noise in cells. To better understand the behaviors and possible extensions of existing models, we design an exact stochastic simulation algorithm for a multimolecular transcriptional system with an Ornstein-Uhlenbeck birth rate that is implemented via a special function-based time-stepping algorithm. We demonstrate that its joint copy-number distributions reduce to analytically well-studied cases in several limiting regimes, and suggest avenues for generalizations.

## 2 Background

### 2.1 Markov modeling of transcriptional processes

Recent methods in single-cell transcriptomics have enabled increasingly precise measurements of copies of mRNA molecules in cells [1, 2]. These experimental improvements have dovetailed with theoretical and computational improvements in modeling transcription in cells. The chemical master equation (CME) is the standard modeling framework for discrete-valued processes [3], providing a natural representation of biomolecular counts. CME models that can be used to compute entire discrete distributions [4] have therefore become increasingly relevant for modeling transcription in cells, and are being used for the purpose of statistical inference of underlying biophysical parameters [5].

No canonical choice of model exists, but certain conventions and assumptions are common (See Figure 1). In the CME formalism, a cell is generally represented as a continuous-time Markov chain traversing a discrete state space. The transitions between states are determined by a set of rates. If all rates are time-independent, residence time in each state is described by an exponential random variable parameterized by the sum of efflux rates, while the choice of transition is made based on the respective relative rates. Furthermore, the state space of the Markov chain may

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

be multi-dimensional and a state may therefore correspond to a *vector* of quantities. A common modeling choice uses $\mathbb{N}_0$ as the domain for dimensions representing molecular species and a finite set of integers as the domain for dimensions representing gene states [6].

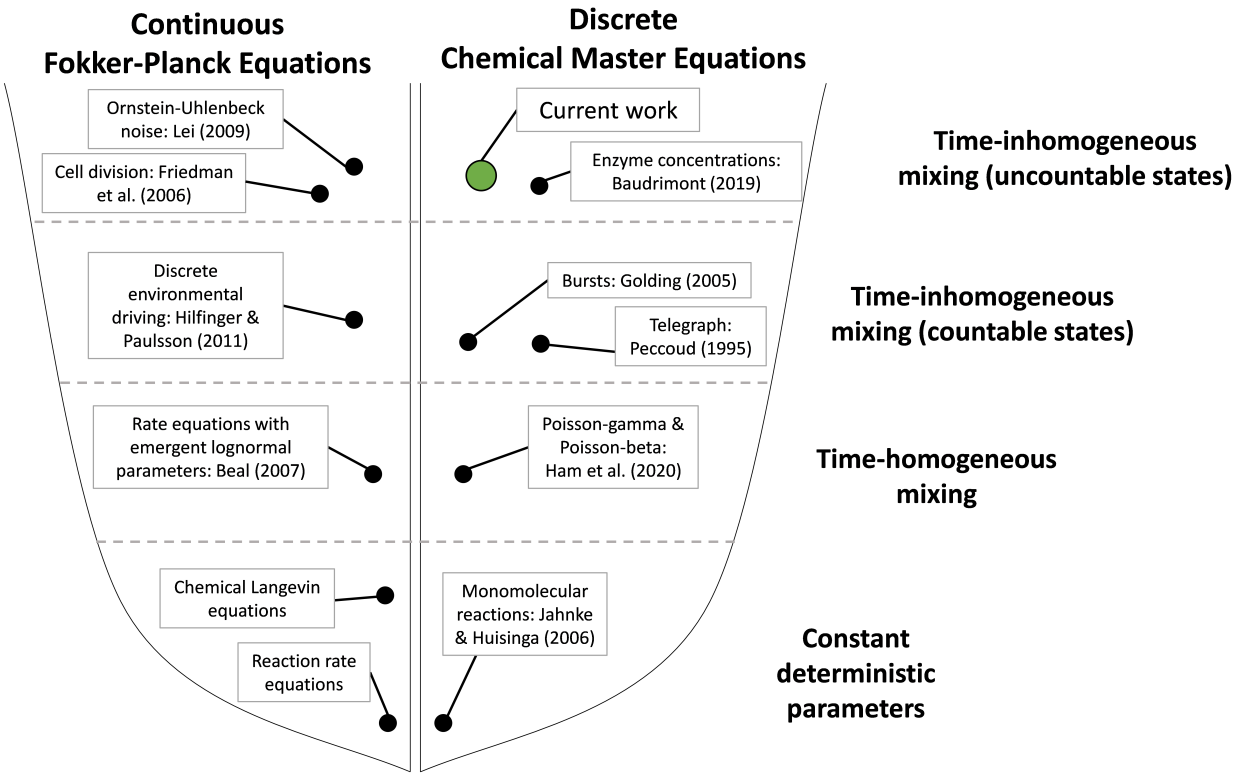


Figure 1: The space of commonly used stochastic models for chemical reactions [2, 6–13].

Given this formalization, it is possible to construct models of gene expression representing physiology of interest. In accordance with recent advances in multimodal data acquisition and modeling [14, 15], we represent mRNA dynamics by a two-stage birth-death process (BDP). A gene locus generates nascent mRNA (unspliced or pre-mRNA) by some variant of a Poisson process. After an exponentially distributed delay with rate $\beta$, nascent mRNA is converted to mature mRNA (spliced mRNA). Finally, after an exponentially distributed delay with rate $\gamma$, the mature mRNA is degraded. The choice of a two-stage BDP is informed by the physiological relevance of splicing and buffering models, as well as our recent discussion of qualitative differences in stationary distributions under different noise models [16].

The simplest stochastic description of gene expression, corresponding to unregulated *constitutive* production, has a single gene state [7] that produces mRNA at a constant rate (Figure 2a). In the light of activation and deactivation known to occur in many biological systems, a more physiologically relevant model posits two distinct gene states with different mRNA production rates, their switching governed by a telegraph process [6]. A common, and physiologically borne out simplification known as the *bursty approximation* [17] considers the limit of infinitesimally short active periods, which generate finite numbers of gene products [2, 18] (Figure 2b). This description produces statistical over-dispersion over the constitutive production model, an excess of variance

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that is referred to as *intrinsic* noise, i.e., noise resulting from non-trivial gene locus dynamics [19]. Another model describes *extrinsic* noise, i.e., noise resulting from cell-to-cell differences in rate parameters (Figure 2c). These two sources have been studied simultaneously since the early 2000s [19,20]; however, full analytical solutions have been limited to rather simple cases due to substantial mathematical complexity.

In the broader context of Figure 1, the more complex transcriptional models described above arise in a natural way from the simpler models by loosening assumptions and replacing constants by stochastic processes. For example, the switching model is equivalent to the birth-death process with a production rate given by the two-state telegraph process, whereas time-homogeneous extrinsic noise models arise by replacing a parameter with a static (non-changing) process initialized at a random value.

Finally, although we describe a method for the simulation of an SDE-driven CME in the current work, there is a wealth of literature using continuous models of gene expressions, essentially treating concentrations rather than single-molecule counts. A standard result in the field [21] describes an equivalence between these two approaches: given a discrete CME, a continuous Fokker-Planck equation (FPE) can be recovered via generating functions; given an FPE, the solution to a CME immediately follows via Poisson mixing. This connection is frequently exploited to make CME systems tractable [22].

## 2.2 Motivation

A recent report considers the extrinsic noise description in the discrete framework [9]. The authors use the gamma distribution to model the transcription rate distribution and motivate the choice by previous results in protein production modeling [11]. The line of reasoning posits that a set of stochastic processes induces a stationary gamma law; therefore, the stationary behavior of a CME under extrinsic noise can be simply computed as a mixture distribution. Specifically the molecule copy numbers are governed by a heterogeneous birth-death process, the stationary distribution is Poisson [7]; if the Poisson rate is, in turn, gamma-distributed, the mixed stationary distribution is negative binomial. The motivation is well in line with previous work [8], which uses the Central Limit Theorem to explain the log-normal distribution of parameter values as a natural consequence of multiplicative effects.

However, upon inspection, this description raises further questions. Although an explanatory process is invoked to motivate the choice of distribution, the process dynamics are disregarded. This is an acceptable approximation in the limit of extremely large time-scale separation – such that the process is substantially slower than the mRNA dynamics, leading to local equilibration of the processes downstream of the gene locus – but it cannot be expected to hold in all timescale regimes. Therefore, we seek to investigate the behavior of CME systems under stochastic driving at the gene locus. To exactly match the asymptotically slow extrinsic noise regime, the process $K(t)$ describing the transcription rate must have a gamma stationary distribution; to reflect transient dynamics, it must also have nontrivial trajectories. The natural representation is a mean-reverting process described by a stochastic differential equation (SDE). Therefore, the problem requires the coupling of a CME to an SDE, which is rather nontrivial. Although analogous problems have been explored in the domain of multi-scale modeling over the past twenty years [23,24], analytical solutions are rare, and none appear to be available for this model. Even simulation is challenging: standard methods tend to use models with Brownian noise in the SDE, and require Euler–Maruyama stochastic integration combined with various rejection schema [24–26]. Intuitively, these models not amenable

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to *direct* (non-rejection) simulation [27] because the assumption of time-independent rates is broken, and the residence time must be computed using numerical integration.

To side-step this problem, we use the gamma Ornstein-Uhlenbeck (Γ-OU) model, which is well-known from quantitative finance [28] and has recently been applied in an investigation of intrinsic noise, albeit with rather different gene dynamics [18]. The Γ-OU model, which does not have a Brownian noise component, affords a semi-analytical solution for the state residence time, and thus enables simulation through a variant of Gillespie's direct method [27]. We describe an algorithm to compute exact residence times, discuss several points pertaining to efficient numerical implementations, and demonstrate that the algorithm is capable of recapitulating the intrinsic, extrinsic, and constitutive models in several degenerate parameter regimes.

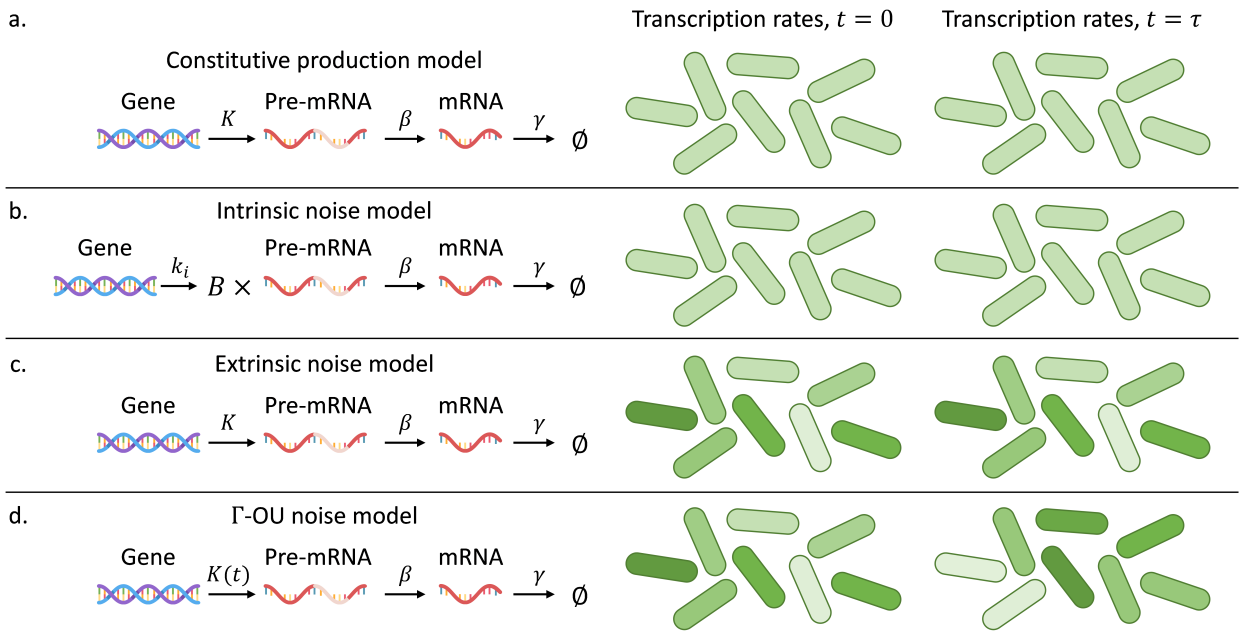


Figure 2: (a) Schema of the intrinsic noise model ($k_i$: burst frequency; $B$: burst size drawn from a geometric distribution; $\beta$: pre-mRNA splicing rate; $\gamma$: mRNA degradation rate. Uniform shade of green indicates identical parameter values across all cells). (b) Schema of the extrinsic noise model ($K$: transcription rate; $\beta$: pre-mRNA splicing rate; $\gamma$: mRNA degradation rate. Different shades of green indicate different, but time-independent, values of $K$ across cells). (c) Schema of the constitutive production model ($K$: transcription rate; $\beta$: pre-mRNA splicing rate; $\gamma$: mRNA degradation rate. Uniform shade of green indicates identical parameter values across all cells). (d) Schema of the Γ-OU noise model ($K(t)$: transcription rate; $\beta$: pre-mRNA splicing rate; $\gamma$: mRNA degradation rate. Different shades of green indicate different values of $K$ across cells and throughout time).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# 3 Notation

## 3.1 Probability distributions

The continuous uniform distribution is represented by $X \sim U(a,b)$, $f(x;a,b) = \frac{1}{b-a}$, where $x \in [a,b]$ and $a,b \in \mathbb{R}$. The geometric distribution is represented by $X \sim Geom(p)$, $P(X=k;p) = (1-p)^k p$, where $k \in \mathbb{N}_0$ and $p \in (0,1]$. The geometric distribution is well-known to arise in the short-burst limit of the two-state transcription model [2]. The negative binomial distribution is represented by $X \sim NegBin(r,p)$, $P(X=k;r,p) = \frac{\Gamma(r+k)}{k!\Gamma(r)}(1-p)^r p^k$, where $k \in \mathbb{N}_0$, $p \in [0,1]$, and $r > 0$. We note that MATLAB and the NumPy library use the opposite convention, with a $\tilde{p}$ parameter defined as $1-p$. The exponential distribution is represented by $X \sim Exp(\eta)$, $f(x;\eta) = \eta e^{-\eta x}$, where $x, \eta > 0$. This is the rate parametrization. We note that MATLAB and the NumPy library use the inverse scale parametrization with parameter $\theta = \eta^{-1}$. The gamma distribution is represented by $X \sim Gamma(\alpha,\eta)$, $f(x;\alpha,\eta) = \frac{\eta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\eta x}$, where $x, \alpha, \eta > 0$. This is the shape/rate parametrization. We note that MATLAB and the NumPy library use the inverse shape/scale parametrization with parameter $\theta = \eta^{-1}$. In the literature, the rate $\eta$ is usually given the variable name "$\beta$"; however, we use the current convention to prevent confusion with the splicing rate parameter. $Exp(\eta)$ is equivalent to $Gamma(1,\eta)$.

## 3.2 Stochastic processes

We follow the mathematical finance convention for the Γ-OU process [18,29]. Specifically, a generalized OU process $K(t)$ is the solution of the SDE

$$dK(t) = -\kappa K(t)dt + dZ(t),$$

where $\kappa > 0$, $K(0) = K_0$ $P$-almost surely, and $Z$ is a subordinator of choice [30]. The Γ-OU process uses the compound Poisson subordinator $Z(t) = \sum_{k=0}^{N_P(t)} J_k$, where $N_P(t)$ is a Poisson counting process with rate $\lambda$, and independent random jump sizes $J_k \sim Exp(\beta)$. The previously reported solution [30] yields

$$K(t) = \sum_{k=0}^{N_P(t)} e^{-\kappa(t-\tau_k)} J_k,$$

where $\tau_k$ are the jump times of $N_P$. Note that $J_0 := K_0$ and $\tau_0 := 0$. The resulting stationary distribution is $Gamma(\frac{\lambda}{\kappa}, \beta)$.

# 4 Simulation design

## 4.1 Simulation of the Γ-OU process

We consider the standard case of simulation on $t \in [0,T]$. The number of Poisson arrivals in this interval follows from the definition of a Poisson process: $N_P(T) \sim Poisson(\lambda T)$. It is well-known [31] that the arrival times of a Poisson counting process on $t \in [0,T]$ are identically distributed to the rank statistics of a uniformly distributed random variable. Therefore, given $N_P(T)$ total

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

jumps, their times $\tau_k, k > 0$ can be computed by drawing $N_P(T)$ random numbers from $U(0, T)$ and sorting the resulting values. The jump sizes $J_k, k > 0$ are computed by drawing $N_P(T)$ exponential random variables with rate $\eta$ or mean $\eta^{-1}$. Given an initial condition, the total number of jumps, their arrival times, and their magnitudes, the Γ-OU process path is fully determined and can be easily computed.

### 4.2 Simulation of the CME

We consider a birth-death system with a single time-inhomogeneous birth rate. For illustration, we consider *n*ascent and *m*ature mRNA species, with respective instantaneous counts $n_n$ and $n_m$. Specifically, we consider three reactions: production with rate $a_1 = K(t)$, splicing with overall rate $a_2 = \beta n_n$, and degradation with overall rate $a_3 = \gamma n_m$. Extensions to more general schema for processing downstream of transcription are trivial. The algorithm is outlined below. The full derivation, including the formula for $\tau$ at each step and numerical considerations for implementation, is provided in Section S2.

1. Set $t = 0$. Initialize $n_n$ and $n_m$.
2. Generate two uniform random variables $u_1$ and $u_2$.
3. Compute time step $\tau$ that meets the criterion $\tau(\beta n_n + \gamma n_m) + \int_t^{t+\tau} K(t')dt' = g(\tau) = \ln(1/u_1)$.
   (a) Check whether the criterion $g(\tau) > \ln(1/u_1)$ at the next jump in transcription rate:
      i. if so, use the Lambert $W$ function to explicitly compute $\tau$,
      ii. If not, check the next jump.
4. Compute instantaneous reaction rates $a_\mu$, $\mu \in \{1, 2, 3\}$.
5. Compute net state efflux rate $a = \sum_{\mu=1}^{3} a_\mu$.
6. Select reaction index $\mu$ to be the lowest $i$ such that $\sum_{\nu=1}^{i} a_\nu > u_2 a$.
7. Advance time by $\tau$.
8. Modify state variables according to the value of $\mu$:
   8.1. $\mu = 1$, $n_n \leftarrow n_n + 1$
   8.2. $\mu = 2$, $n_n \leftarrow n_n - 1$, $n_m \leftarrow n_m + 1$
   8.3. $\mu = 3$, $n_m \leftarrow n_m - 1$
9. Return to step 2.

## 5 Results

### 5.1 Asymptotic regimes

In the current section, we consider several asymptotic regimes where the SDE-driven model reduces to common physiological models of transcription. In all regimes, we simulate $10^4$ cells and use the downstream process rate parameters $\beta = 1.2$ and $\gamma = 0.7$. The stationary distributions are shown in Figure 3.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

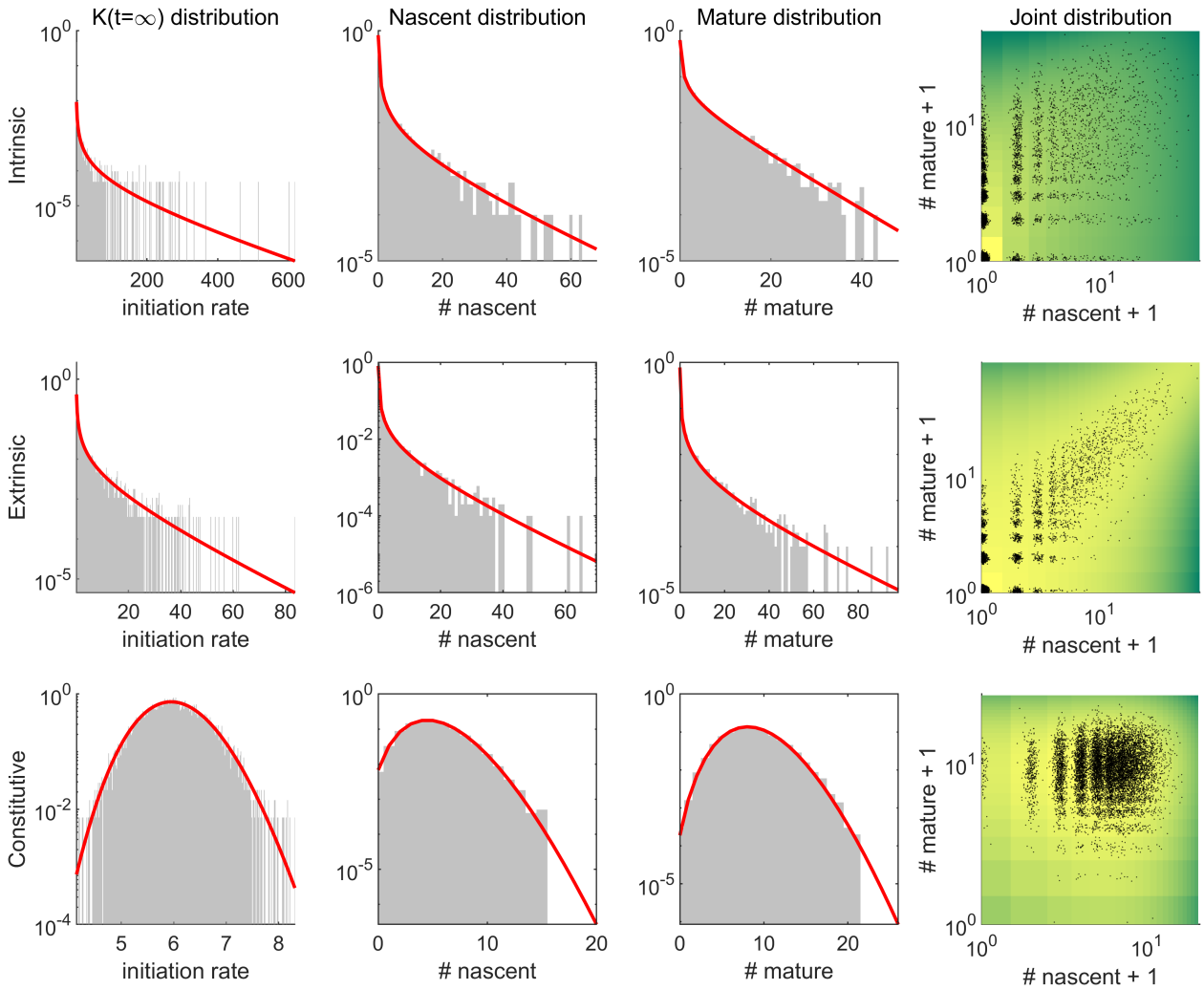


Figure 3: Simulation results in three asymptotic regimes (grey histograms: observed distributions; red lines: analytical results; black points: cells; right row color: log analytical joint PMF). [Code]

### 5.1.1 Intrinsic-only noise

As $\kappa \to \infty$ and $\eta \to 0$, the Γ-OU stochastic process reduces to a series of peaks of infinitesimally short duration. If $\kappa\eta \to b$, a finite quantity, the mass of each peak is finite and given by $J_i/\kappa$. This case reduces to the bursty system studied by Amrhein et al. [18], with burst arrival rate $\lambda$ and mean burst size $b$. The agreement in this domain is shown in the first row of Figure 3. The time series is provided in Figure S1. The parameters used for the simulation are $\kappa = 10$, $\lambda = 0.1$, and $\eta = 6.7 \times 10^{-3}$, with effective mean burst size $b = 15$.

### 5.1.2 Extrinsic-only noise

Purely extrinsic noise is conventionally modeled as a mixture with time-independent, gamma-distributed transcription parameters. In the context of transcription governed by a Γ-OU process, this corresponds, intuitively enough, to a regime with significant timescale separation between the

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

gene locus noise and the downstream processing. Specifically, if $\kappa \ll \beta, \gamma$, the cells experience local equilibrium. To yield a non-degenerate stationary distribution of rates, $\lambda$ must also vanish. Therefore, extrinsic noise is recapitulated whenever SDE dynamics are sparse and slow compared to downstream kinetics. The agreement in this domain is shown in the second row of Figure 3. The time series is provided in Figure S2. The parameters used for the simulation are $\kappa = 0.12$, $\lambda = 0.01$, and $\eta = 6.7 \times 10^{-2}$.

#### 5.1.3 Constitutive production

The stationary distribution rates are distributed per $Gamma(\lambda/\kappa, \eta)$, with mean $\frac{\lambda}{\kappa\eta}$ and variance $\frac{\lambda}{\kappa\eta^2}$. Therefore, if $\eta \to \infty$ while $\lambda/\kappa \to \infty$, with the finite constraint $\frac{\lambda}{\kappa\eta} = \mu$, the stationary distribution reduces to a Dirac $\delta$ distribution with a point mass at $\mu$. This degenerate case yields constitutive production with a constant transcription rate $\mu$. The downstream nascent and mature mRNA distributions are given by $N \sim Poisson(\mu/\beta)$ and $M \sim Poisson(\mu/\gamma)$, in accord with standard results [7]. The agreement in this domain is shown in the third row of Figure 3. The time series is provided in Figure S3. The parameters used for the simulation are $\kappa = 8.3 \times 10^{-4}$, $\lambda = 0.1$, and $\eta = 20$, with effective mean initiation rate $\mu = 6$.

### 5.2 General parameter regime

The algorithm permits the exact, direct simulation of the SDE-driven system with arbitrary dynamics. The results of a sample simulation with comparable rates are shown in Figure 3. The transcriptional rate agrees with the intended theoretical form, both throughout the time series and at steady state. Furthermore, the observed long-term nascent and mature mRNA means agree with the theoretical stationary expectations $\mathbb{E}[K]/\beta$ and $\mathbb{E}[K]/\gamma$. We do not derive expressions for the joint or marginal mRNA distributions, but note that the marginals agree fairly well with a negative binomial fit.

## 6 Discussion

We have developed an exact and direct routine to simulate an SDE coupled to a birth-death process and discussed its reduction to a set of qualitatively distinct and physiologically relevant gene expression regimes, as shown in Figure 3. Furthermore, it is suitable for simulations in intermediate parameter regimes, with a sample simulation depicted in Figure 4. In the current section, we suggest extensions to broader classes of stochastic processes, as well as theoretical directions and implications.

### 6.1 Generalizations to a broader class of subordinators

Due our focus on stationary behavior under gamma-distributed transcriptional parameters, we only explore a fairly limited domain of $K(t)$ behaviors. However, as evident from the functional form, transcriptional dynamics driven by *any* subordinator $Z(t)$ containing a finite number of jump in every closed interval can be simulated using an identical procedure. Specifically, the core loop does not depend on the details of the random number generation process that produces $N_P(t)$, $\tau_k$, and $J_k$. Therefore, a fairly wide range of $K(t)$ dynamics can be pursued with minimal modifications to the algorithm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

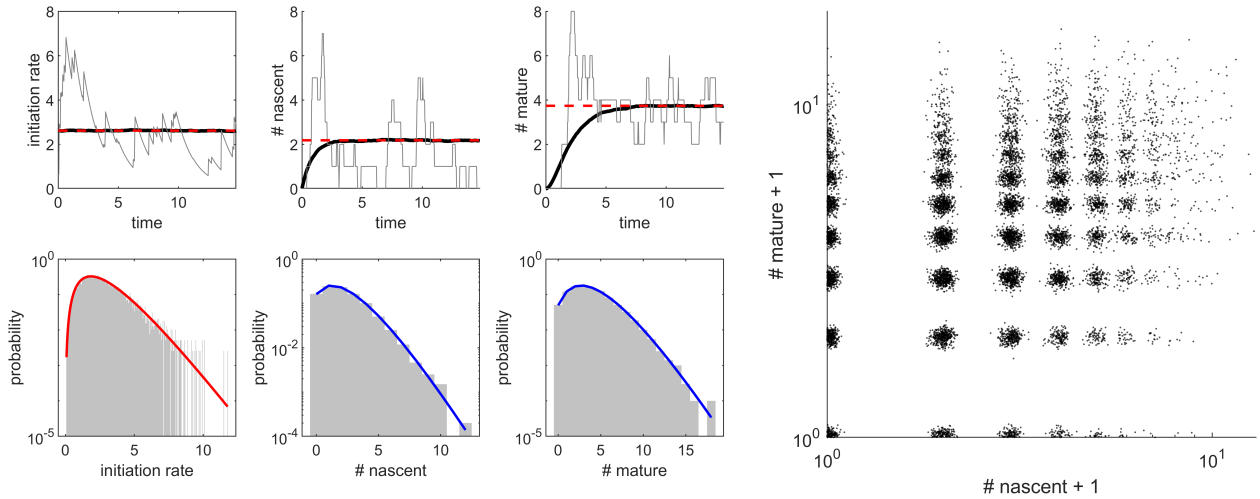


Figure 4: Sample simulation results ($\kappa = 0.6765$, $\lambda = 2.3$, $\eta = 1.3$, $\beta = 1.2$, $\gamma = 0.7$, $10^4$ cells). Top row, left: time series for initiation rate, number of nascent mRNA, number of mature mRNA (black line: mean of all iterations; grey line: single iteration, red dashed line: expected stationary mean). Bottom row, left: stationary distributions (grey histogram: observed distribution; red line: expected analytical distribution; blue line: best negative binomial fit). Right: empirical joint distribution (black points: cells; normal jitter with $\sigma = 0.05$ added).

The case of time-varying $\eta$ is trivial to implement. Given a function which defines $\eta(t)$, it is only necessary to draw $N_P(T)$ exponential random variables with rates $\eta(t_1), \eta(\tau_2), ..., \eta(\tau_{N_P(T)})$. The resulting simulation is exact for analytical $\eta(t)$. It may be approximate for more complex forms of $\eta(t)$ given by, e.g., deterministic or stochastic differential equations. The case of time-varying $\lambda$ is likewise straightforward; procedures for the simulation of inhomogeneous Poisson processes are readily available [32]. Parenthetically, we note that if $\lambda(t)$ has an analytical integral, its simulation is equivalent to the exact Gillespie simulation of a pure-birth CME system; therefore, methods described elsewhere in the report can be used to exactly simulate the process arrival times.
The simulation is by no means limited to the simple two-step BDP illustrated here. In fact, the form of the root-finding problem in $\tau$ that gives rise to Equation S1 naturally suggests that the birth process can be combined with any number of time-homogeneous downstream processes. Therefore, the framework is immediately applicable to arbitrary downstream reaction graphs.
By the thinning property of a Poisson process, the simulation of a single gene locus with arrival rate $\lambda$ is equivalent to the simultaneous simulation of $n$ gene loci with rates $\lambda_1, \lambda_2, ..., \lambda_n$. The event–locus assignments are performed by drawing from a categorical distribution (with PMF $p_i = \lambda_i/\lambda$). Finally, given different rates $\eta_i$ at each locus, the overall compound Poisson process is produced by drawing from the distribution $Exp(\eta_i)$ at each event $i$.

## 6.2 Theoretical directions

Our unified description of noise sources presents an alternative to phenomenological additive or multiplicative noise models [10]. Further appeal lies in the functional form of the Γ-OU model, which permits interpretation in terms of site exposure followed by linear occlusion.
We suggest that the intermediate regime of SDE parameters, with moderate rates $\kappa$ and $\lambda$, presents

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a natural area for exploration, as it interpolates between the intrinsic and extrinsic noise regimes. Qualitatively, the intrinsic noise regime corresponds to dispersion purely governed by the spike masses and arrivals, whereas the extrinsic noise corresponds to dispersion purely governed by spike magnitudes and the ratio of $K(t)$ rates. We speculate that the intermediate regime corresponds to dispersion greater than either extreme allows, and thus permits the integration of both noise models in a single, self-consistent mechanistic description.

Although specific analytical solutions are outside of the scope of the current investigation, we suggest that the apparent agreement in Figure 4 (blue lines with best negative binomial fits) is only approximate, at least for the mature marginal. This hypothesis is based on the imperfect agreement between the mature marginal and the negative binomial distribution in the limit of pure intrinsic noise [33].

Throughout the current work, we focus upon the two-stage BDP. As recently discussed [16], the use of a multi-stage model yields strikingly discordant results under the assumptions of pure intrinsic and extrinsic noise. Conversely, the availability of multimodal data presents opportunities for model discrimination, even at steady state. Therefore, we suggest that multimodal information may be highly informative in intermediate regimes.

## 7 Code availability

MATLAB code that can be used to reproduce Figures 4-S3, including the simulation and plotting routines, is available at `https://github.com/pachterlab/GP_2021`.

## 8 Acknowledgments

The DNA, pre-mRNA, and mature mRNA illustrations used in Figure 2, reproduced from [34], are derivatives of the DNA Twemoji by Twitter, Inc., used under CC-BY 4.0. G.G. and L.P. were partially funded by NIH U19MH114830.

## References

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

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

[3] David Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York, 1987.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[4] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124(4):044104, 2006.

[5] Brian Munsky, Guoliang Li, Zachary R. Fox, Douglas P. Shepherd, and Gregor Neuert. Distribution shapes govern the discovery of predictive models for gene regulation. *Proceedings of the National Academy of Sciences*, 115(29):7533–7538, 2018.

[6] Jean Peccoud and Bernard Ycard. Markovian Modeling of Gene Product Synthesis. *Theoretical Population Biology*, 48(2):222–234, 1995.

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

[8] Jacob Beal. Biochemical complexity drives log-normal variation in genetic expression. *Engineering Biology*, 1(1):55–60, June 2017.

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

[10] A. Hilfinger and J. Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences*, 108(29):12167–12172, July 2011.

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

[12] Jinzhi Lei. Stochasticity in single gene expression with both intrinsic noise and fluctuation in kinetic parameters. *Journal of Theoretical Biology*, 256(4):485–492, February 2009.

[13] Antoine Baudrimont, Vincent Jaquet, Sandrine Wallerich, Sylvia Voegeli, and Attila Becskei. Contribution of RNA Degradation to Intrinsic and Extrinsic Noise in Gene Expression. *Cell Reports*, 26(13):3752–3761.e5, March 2019.

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

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

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

[17] 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

modulated across the human genome. *Proceedings of the National Academy of Sciences*, 109(43):17454–17459, October 2012.

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

[19] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, October 2002.

[20] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584):1183–1186, 2002.

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

[22] V. Shahrezaei and P. S. Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, 105(45):17256–17261, November 2008.

[23] Eric L. Haseltine and James B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *The Journal of Chemical Physics*, 117(15):6959–6969, October 2002.

[24] J. Pahle. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches. *Briefings in Bioinformatics*, 10(1):53–64, October 2008.

[25] Martin Bentele and Roland Eils. General Stochastic Hybrid Method for the Simulation of Chemical Reaction Processes in Cells. In Vincent Danos and Vincent Schachter, editors, *Computational Methods in Systems Biology*, volume 3082, pages 248–251. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. Series Title: Lecture Notes in Computer Science.

[26] Howard Salis and Yiannis Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *The Journal of Chemical Physics*, 122(5):054103, February 2005.

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

[28] Wim Schoutens. *Lévy Processes in Finance*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, Chichester, UK, March 2003.

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

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

[31] D.R. Cox and H.D Miller. *The Theory of Stochastic Processes*. Chapman & Hall, 2001.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[32] P. A. W Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, September 1979.

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

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

[35] A. Prados, J. J. Brey, and B. Sánchez-Rey. A dynamical monte carlo algorithm for master equations with time-dependent transition rates. *Journal of Statistical Physics*, 89(3-4):709–734, November 1997.

[36] Roberto Iacono and John P. Boyd. New approximations to the principal real-valued branch of the Lambert W-function. *Advances in Computational Mathematics*, 43(6):1403–1436, December 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# S1 Supplementary figures

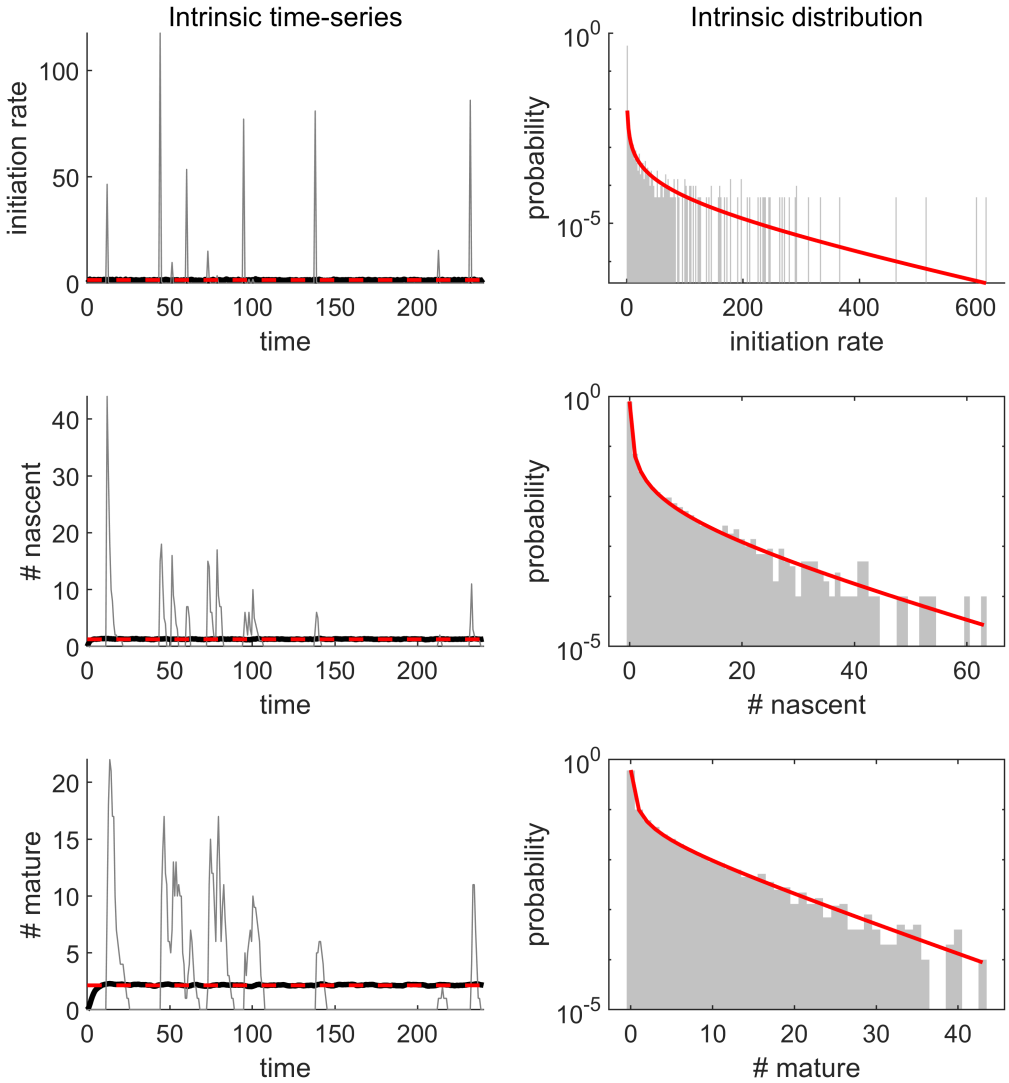


Figure S1: Simulation results in the intrinsic noise regime. Left: time series for initiation rate, number of nascent mRNA, number of mature mRNA (black line: mean of all iterations; grey line: single iteration, red dashed line: expected stationary mean). Right: stationary distributions (grey histogram: observed distribution; red line: expected analytical distribution).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

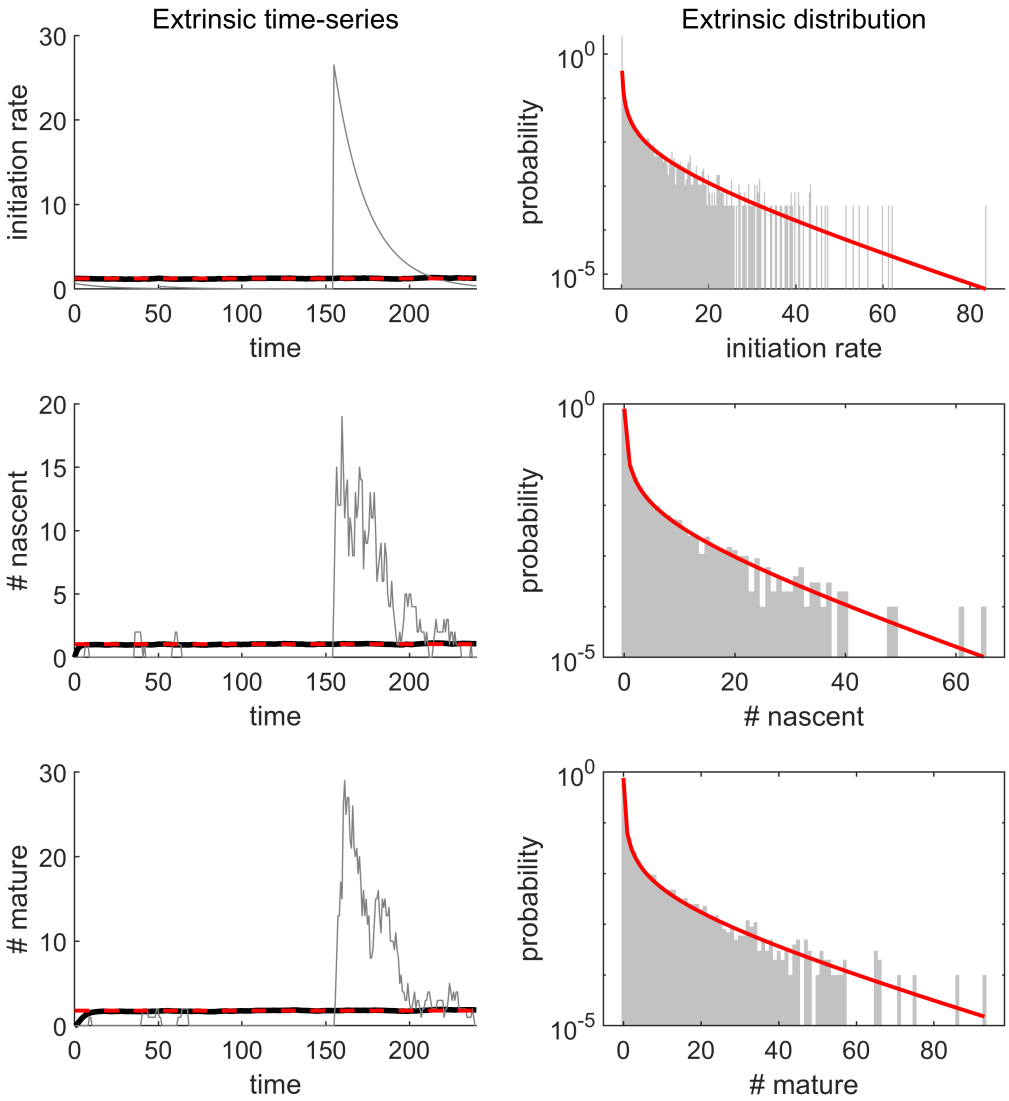


Figure S2: Simulation results in the extrinsic noise regime. Left: time series for initiation rate, number of nascent mRNA, number of mature mRNA (black line: mean of all iterations; grey line: single iteration, red dashed line: expected stationary mean). Right: stationary distributions (grey histogram: observed distribution; red line: expected analytical distribution).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

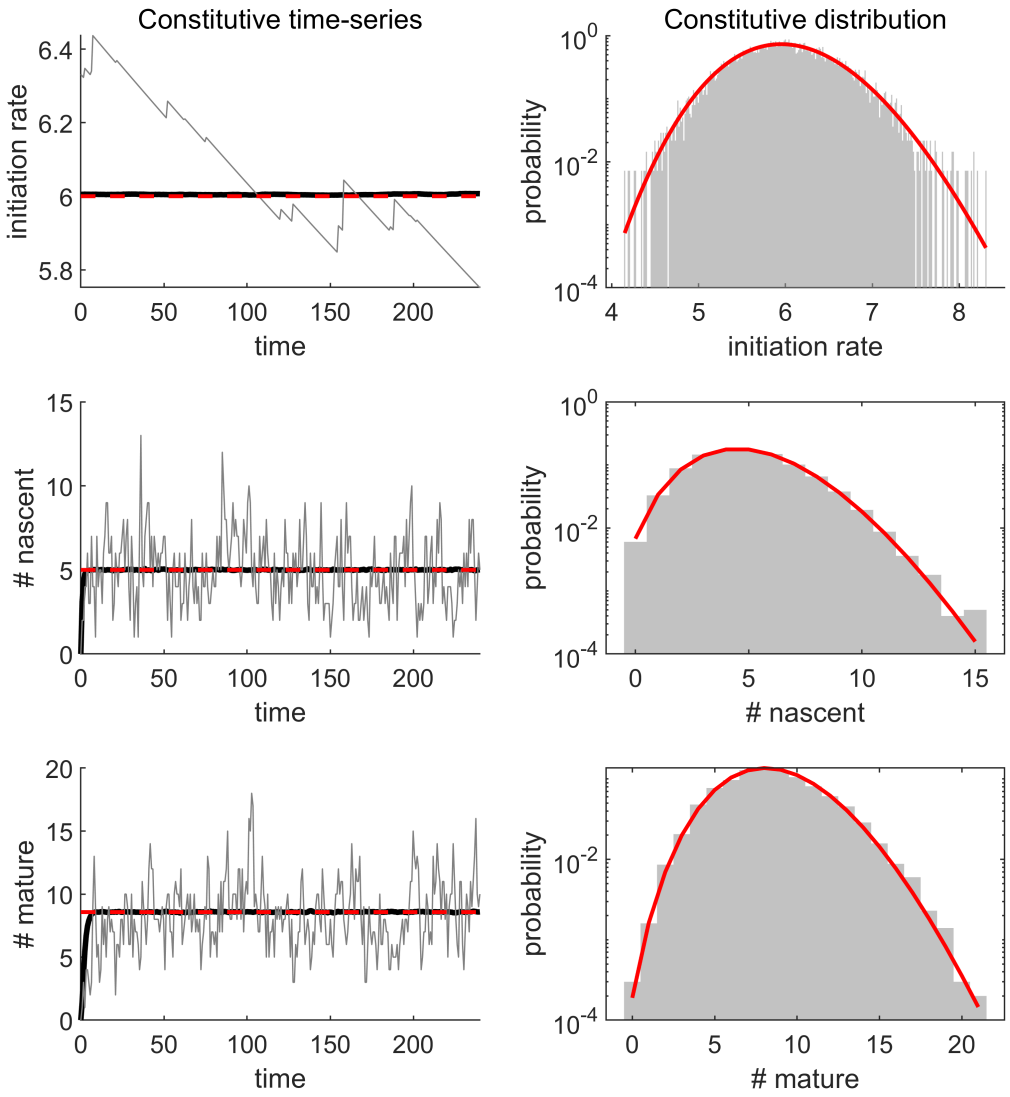


Figure S3: Simulation results in the constitutive production regime. Left: time series for initiation rate, number of nascent mRNA, number of mature mRNA (black line: mean of all iterations; grey line: single iteration, red dashed line: expected stationary mean). Right: stationary distributions (grey histogram: observed distribution; red line: expected analytical distribution).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# S2 Supplementary information

## S2.1 Time-homogeneous algorithm

For comparison, the standard time-homogeneous stochastic simulation algorithm (Gillespie algorithm) proceeds as follows [27]:

1. Set $t = 0$. Initialize $n_n$ and $n_m$.
2. Compute instantaneous reaction rates $a_\mu$, $\mu \in \{1, 2, 3\}$.
3. Compute net state efflux rate $a = \sum_{\mu=1}^{3} a_\mu$.
4. Generate two uniform random variables $u_1$ and $u_2$.
5. Compute time step $\tau = \frac{1}{a} \ln(1/u_1)$.
6. Select reaction index $\mu$ such that $\sum_{\nu=1}^{\mu-1} a_\nu < u_2 a \leq \sum_{\nu=1}^{\mu} a_\nu$.
7. Advance time by $\tau$.
8. Modify state variables according to the value of $\mu$:
   - 8.1. $\mu = 1$, $n_n \leftarrow n_n + 1$
   - 8.2. $\mu = 2$, $n_n \leftarrow n_n - 1$, $n_m \leftarrow n_m + 1$
   - 8.3. $\mu = 3$, $n_m \leftarrow n_m - 1$
9. Return to step 2.

Since the computation of $\tau$ presupposes constant reaction rates, this algorithm is inappropriate for the time-inhomogeneous case.

## S2.2 Time-inhomogeneous algorithm

The case of a time-inhomogeneous birth rate necessitates a more complex coupled computation [35]. Specifically, the random time step $\tau$ is selected according to $\int_t^{t+\tau} a(t')dt' = \ln(1/u_1) = \Lambda$. Using the definition of $a$:

$$\begin{aligned}
\int_t^{t+\tau} a(t')dt' &= \int_t^{t+\tau} \sum_{\mu=1}^{3} a_\mu(t')dt' \\
&= \int_t^{t+\tau} \big(K(t') + \beta n_n + \gamma n_m\big)dt' \\
&= \tau(\beta n_n + \gamma n_m) + \int_t^{t+\tau} K(t')dt'
\end{aligned}$$

Given a particular realization, we can directly integrate $K$. Specifically:

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\int_t^{t+\tau} K(t')dt' = \frac{1}{\kappa}\sum_{k=0}^{N_P(t)} e^{-\kappa(t-\tau_k)}J_k - \frac{1}{\kappa}\sum_{k=0}^{N_P(t+\tau)} e^{-\kappa(t+\tau-\tau_k)}J_k$$

This quantity is straightforward to evaluate. However, the specific functional form makes it challenging to compute $\tau$ without resorting to numerical root-finding algorithms. Therefore, an alternative approach is desired for fast computation.

We begin by treating the simplest case. If $t > \tau_k$ for all $k$, no more jumps occur after the current time, and $K(t+\tau)$ exponentially decays as a function of $\tau$, with the functional form $K(t+\tau) = K(t)e^{-\kappa\tau}$. Therefore,

$$\begin{aligned}&\tau(\beta n_n + \gamma n_m) + \int_t^{t+\tau} K(t')dt' \\ &= \tau(\beta n_n + \gamma n_m) + \frac{K(t)}{\kappa}(1 - e^{-\kappa\tau})\end{aligned}$$

This implies the root-finding problem in $\tau$:

$$\begin{aligned}\Lambda &= \tau(\beta n_n + \gamma n_m) + \frac{K(t)}{\kappa}(1 - e^{-\kappa\tau}) \\ 0 &= \tau(\beta n_n + \gamma n_m) - \frac{K(t)}{\kappa}e^{-\kappa\tau} + \left(\frac{K(t)}{\kappa} - \Lambda\right) \\ 0 &= C_1\tau - C_2(t)e^{-\kappa\tau} + C_3(t)\end{aligned}$$

This equation has the analytical solution:

$$\tau = \frac{1}{\kappa}W\left(\frac{\kappa C_2}{C_1}e^{\kappa C_3/C_1}\right) - \frac{C_3}{C_1} = \phi_W(t) \tag{S1}$$

where $C_1, C_2$, and $C_3$ are evaluated at $t$, whereas $W$ is the product logarithm function, i.e. $W_0$, the principal branch of the Lambert $W$ function. This solution is straightforward to compute using standard packages, such as the MATLAB Symbolic Toolbox and the SciPy library for Python. The alternative formulation is relevant when $C_1 = 0$:

$$\tau = -\frac{1}{\kappa}\ln\left(\frac{C_3(t)}{C_2(t)}\right) \tag{S2}$$

Parenthetically, we note the terminal case $t+\tau > T$, i.e. that the reaction flux up to $T$ is insufficient to match $\Lambda$. Although the SDE dynamics are not simulated past $T$, and no information about $K$ is known past this time horizon, this is not a problem; the simulation remains exact up until $T$, where it halts. Another edge case, where $\phi_W(t)$ is complex-valued, implies that the total reaction flux up to $t = \infty$ is insufficient to meet $\Lambda$, and again simply leads to the termination of the simulation at

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$T$. This edge case only occurs when $C_1 = 0$, as the downstream reactions occur in finite time in the converse case.

Next, we consider the first non-trivial extension: $t < \tau_k$ for a single $k$; a single jump occurs after the current time. For convenience of notation, we define $\tau_N := \tau_{N_P(T)}$. It remains to bound $t + \tau$ within the region $(t, \tau_N)$ or the region $(\tau_N, \infty)$.

Since $g(\tau; t) = \tau(\beta n_n + \gamma n_m) + \int_t^{t+\tau} K(t')dt'$ is guaranteed to be monotonic, we can use a simple binary decision procedure. If $g(\tau_N - t; t) = (\tau_N - t)(\beta n_n + \gamma n_m) + \frac{K(t)}{\kappa}(1 - e^{-\kappa(\tau_N - t)}) > \Lambda$, the value of the integral up to $\tau_N$ is an overestimate and the solution is given by Equation S1 evaluated at $t$, i.e. $\phi_W(t)$. If the converse is true, the value is an underestimate and the solution is given by $\phi_W(\tau_N) + (\tau_N - t)$.

This procedure can be extended to an arbitrary number of jumps after $t$. The implementation requires a choice of a search procedure; we choose a simple rightward scan. Specifically, given $t < \tau_k < \tau_{k+1} < ... < \tau_N$:

1. Assign upper bound for the integral $L \leftarrow k$ and running time $t_R \leftarrow t$.
2. Check whether $L \leq N$.
   - 2.1. If $L \leq N$, evaluate $G = g(\tau_L - t; t) = g(\tau_k - t; t) + g(\tau_L - \tau_k; \tau_k)$.
     - 2.1.1. If $G > \Lambda$, the solution is given by $\phi_W(t_R) + (t_R - t)$.
     - 2.1.2. If $G < \Lambda$, assign $L \leftarrow L + 1$ and $t_R \leftarrow \tau_L$.
     - 2.1.3. Return to 2.
   - 2.2. If $L > N$, the solution is given by $\phi_W(t_R) + (t_R - t)$.

Since $K(t)$ is known, it is trivial to pre-compute the quantities $\int_{\tau_i}^{\tau_{i+1}} K(t')dt'$, $i \in \{0, 1, ..., N_P(T) - 1\}$, where $\tau_0 := 0$. Therefore, computing the term $g(\tau_L - \tau_k; \tau_k)$ requires a summation over the pre-computed integral terms $\sum_{i=k}^{L-1} \int_{\tau_i}^{\tau_{i+1}} K(t')dt'$ and a single evaluation of the exponential-exit product $(\tau_L - \tau_k)(\beta n_n + \gamma n_m)$. Finally, the remainder $g(\tau_k - t; t)$ requires one evaluation of the analytical integral per Gillespie time step.

With $\tau$ determined, it remains to select the specific reaction channel. The exponential-exit weights are given by $a_2 = \tau\beta n_n$ and $a_3 = \tau\gamma n_m$. The weight $a_1$ of the birth reaction is given by $\int_t^{t+\tau} K(t')dt'$, which is given by

$$\frac{K(t)}{\kappa}(1 - e^{-\kappa\tau})$$

if no jumps occur up within $(t, t + \tau)$, and

$$\frac{K(t)}{\kappa}(1 - e^{-\kappa(\tau_k - t)}) + \sum_{i=k}^{M-1} \int_{\tau_i}^{\tau_{i+1}} K(t')dt' + \frac{K(\tau_M)}{\kappa}(1 - e^{-\kappa(\tau + t - \tau_M)})$$

if $t < \tau_k < \tau_{k+1} < ... < \tau_M < t + \tau$. Therefore, the following steps of the time-inhomogeneous algorithm are identical to steps 6-9 of the time-homogeneous algorithm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.20.427480; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

### S2.3 Implementation details

Several points regarding the efficient implementation of the algorithm bear further discussion.
For computational facility, at each step of the Gillespie simulation, we set $\tau_{k-1} \to t$ and $K(\tau_{k-1}) \to K(t)$. This approach creates a virtual jump at the current time, and allows treating the integral $\int_t^{\tau_k} K(t')dt'$ without creating a special edge case. Furthermore, to minimize the number of times the pre-computed integrals are accessed, we compute $\Delta G$ at each step, compare it to $\Lambda$, and *decrement* $\Lambda$ by $\Delta G$ if the reaction flux is insufficient.
The formulation in Equation S1 is susceptible to overflow as $\kappa C_3/C_1 \to \infty$. A naïve computation at sufficiently high values yields $e^{\kappa C_3/C_1} = \infty$ and $\tau = \infty$, halting the simulation. Therefore, wherever overflow is likely to occur, it is necessary to use the appropriate approximation to $W$. We follow the approach of Iacono and Boyd [36].
As $x \to \infty$, $\ln(1+x)$ has the Puiseux series representation $\ln(x) + x^{-1} + O(x^{-2})$. For $x$ sufficiently high to produce overflow, we truncate at the first term and use $\ln(1+x) \approx \ln(x)$.
As an initial guess, we can choose $W_0(x) = \ln(1 + x\zeta(x))$, where $\zeta(x) = \frac{1}{1+0.5\ln(1+x)}$; we note that the subscript refers to the approximation order rather than the branch of the function. Using the Puiseux series, $\zeta(x) \approx \frac{1}{1+0.5\ln x}$. Assuming $x$ is high enough, we can further assume $\ln(1+x\zeta(x)) \approx \ln(x\zeta(x)) = \ln x - \ln(1 + 0.5\ln(x))$. Higher-order approximations follow from the iterative schema $W_{n+1} = \frac{W_n}{1+W_n}(1 + \ln x - \ln W_n)$. We use the fifth-order iterative approximation whenever the argument of the Lambert $W$ function is greater than $10^3$.